Base resolution maps reveal the importance of 5-hydroxymethylcytosine in a human glioblastoma

Raiber, Eun-Ang; Beraldi, Dario; Martínez Cuesta, Sergio; McNulty, Gordon R; Kingsbury, Zoya; Becq, Jennifer; James, Terena; Lopes, Margarida; Allinson, Kieren; Field, Sarah; Humphray, Sean; Santarius, Thomas; Watts, Colin; Bentley, David; Balasubramanian, Shankar

DOI: 10.1038/s41525-017-0007-6

License: Creative Commons: Attribution (CC BY)

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Raiber, E-A, Beraldi, D, Martinez Cuesta, S, McNulty, GR, Kingsbury, Z, Becq, J, James, T, Lopes, M, Allinson, K, Field, S, Humphray, S, Santarius, T, Watts, C, Bentley, D & Balasubramanian, S 2017, 'Base resolution maps reveal the importance of 5-hydroxymethylcytosine in a human glioblastoma', NPJ Genomic Medicine, vol. 2, pp. 6. https://doi.org/10.1038/s41525-017-0007-6

Link to publication on Research at Birmingham portal

Publisher Rights Statement:
Checked for eligibility: 02/07/2019
Raiber, Eun-Ang, et al. "Base resolution maps reveal the importance of 5-hydroxymethylcytosine in a human glioblastoma." NPJ Genomic Medicine 2.1 (2017): 6. https://doi.org/10.1038/s41525-017-0007-6

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- Users may use extracts from the document in line with the concept of ‘fair dealing’ under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBITRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 10. Mar. 2020
Base resolution maps reveal the importance of 5-hydroxymethylcytosine in a human glioblastoma

Eun-Ang Raiber1, Dario Beraldi1, Sergio Martinez Cuesta1, Gordon R. McInroy2, Zoya Kingsbury3, Jennifer Becq3, Terena James3, Margarida Lopes5, Kieren Allinson6, Sarah Field1, Sean Humphray3, Thomas Santarius5, Colin Watts5, David Bentley3 and Shankar Balasubramanian1,2,6

Aberrant genetic and epigenetic variations drive malignant transformation and are hallmarks of cancer. Using PCR-free sample preparation we achieved the first in-depth whole genome (hydroxyl)-methylcytosine, single-base-resolution maps from a glioblastoma tumour/margin sample of a patient. Our data provide new insights into how genetic and epigenetic variations are interrelated. In the tumour, global hypermethylation with a depletion of 5-hydroxymethylcytosine was observed. The majority of single nucleotide variations were identified as cytosine-to-thymine deamination products within CpG context, where cytosine was preferentially methylated in the margin. Notably, we observe that cells neighbouring tumour cells display epigenetic alterations characteristic of the tumour itself although genetically they appear "normal". This shows the potential transfer of epigenetic information between cells that contributes to the intratumour heterogeneity of glioblastoma. Together, our reference (epi)-genome provides a human model system for future studies that aim to explore the link between genetic and epigenetic variations in cancer progression.

npj Genomic Medicine (2017) 2:6; doi:10.1038/s41525-017-0007-6

INTRODUCTION

Genetic and epigenetic alterations to the genome shape the development of human malignancies. The patterns of the DNA methylation mark 5-methylcytosine (5mC) become aberrant in human malignancies and affect cellular functions.1 The recently re-discovered DNA mark 5-hydroxymethylcytosine (5hmC)2, 3 is a functionally important DNA modification, and is an intermediate in the process of active demethylation of 5mC. In cancers, 5hmC patterns undergo considerable changes4 that have been linked to genome instability5, 6 and remodelling of the DNA methylation pattern.7 Previous studies revealed that 5hmC is consistently found at significantly reduced levels in various solid tumours.8–10 Indeed, epigenetic regulators such as DNA methyltransferases (DNMT), ten-eleven-translocation (TET) proteins or isocitrate dehydrogenases (IDH), are crucial for normal and malignant cellular development.11 Very few studies however, have effectively mapped the distribution of 5hmC in normal or cancer tissues. Herein, we present the first single base resolution maps of whole genomes, methylomes, and hydroxymethylomes for matched human glioblastoma and tumour margin samples.

RESULTS

Enhanced (hydroxyl)-methylome sequencing reveals global hypermethylation in tumour with loss of 5hmC

We performed whole genome sequencing at 100× coverage of blood, tumour, and margin samples from a glioblastoma patient (Fig. 1a) using a PCR-free library preparation.12 Total RNA sequencing of all three samples was also performed. We employed oxidative bisulfite sequencing (oxBS-seq)13 and bisulfite sequencing (BS-seq) to generate high-depth (80×) sequence coverage and built single-base resolution maps that distinguished 5mC and 5hmC modifications (Fig. 1b and Supplementary Tables 1 and 2). In the margin sample, we found levels of 50% for 5mC and 20% for 5hmC integrated over all CpGs in the genome, whereas in the tumour, we observed global hypermethylation, with average levels of 60% 5mC and a drastic loss of 5hmC to 1.6% (Fig. 1c).

We then assessed specific regions within the genome that were differentially methylated between tumour and margin and identified substantial hypermethylation in the 3′-UTRs, CpG shelves, exons and introns of the tumour compared to the margin (Fig. 1d). We observed that CpGs with high levels of 5hmC in margin tissue have correspondingly higher levels of 5mC in tumour (Fig. 1e) suggesting that elevated 5mC sites in tumour DNA have arisen through failure to oxidise 5mC to 5hmC.

Gene promoter (hydroxyl)-methylation and transcript levels

To investigate the possible mechanism that could account for the loss of 5hmC, we looked for loss-of-function mutations or epimutations and identified single base resolution (hydroxy)-methylation changes in the promoter in connection to transcript levels of key epigenetic regulators (Fig. 2a). Apparent loss of 5hmC in tumours can occur through loss-of-function mutations in TET
enzymes that oxidise 5mC to 5hmC, or inhibition of TET activity by the oncometabolite beta-hydroxyglutarate generated by mutant IDH1/2. 14-17 IDH mutations are mutually exclusive with mutations in TET, at least in acute myeloid leukaemia. 18 No IDH mutations or loss-of-function mutations in the TET genes were observed in the tumour DNA for this patient. However, we observed hypermethylation at TET2/3 gene promoters with concomitant loss of 5hmC at the same CpGs and a corresponding reduction in TET2/3 expression in the tumour (Fig. 2b). TET2 promoter methylation has previously been observed in low-grade diffuse gliomas lacking IDH1/2 mutations and provides a third mechanism to cause loss in maintenance of 5hmC levels in the tumour. Previous literature has also linked reduced TET function to solid and myeloid malignancies, 5, 19, 20 and suggested a key role for TET in the prevention of cancer by suppressing cell invasion 21 and promoting genome integrity. 5, 6 Our results in the current glioblastoma case are consistent with these ideas.

Transcript levels in margin are related to the levels of both, methylation and hydroxymethylation in promoters (Fig. 2c, d). The transcript levels are low at genes whose promoters have a high level of 5mC but a low level of 5hmC (sector 3, Fig. 2c). When 5hmC is high and 5mC is low (sector 2, Fig. 2c) transcript levels are high. This fits a model where 5mC marks silent promoters, while 5hmC and C mark active or poised promoters (Fig. 2d). 22 Interestingly, even when 5mC is high and 5hmC levels shift from low to high (sectors 3 to 4, Fig. 2c), the expression is increased. Quantitative analysis of mRNA levels revealed that 8141 genes (24%) were differentially expressed (logFC > 2 or logFC < −2) between the margin and tumour (Supplementary Fig 2) with Fig. 2e highlighting expression levels of the top ten most differentially hydroxymethylated promoters.

Patterns of genomic variations

We analysed the mutational landscape by identifying single nucleotide variants (SNVs) in tumour and margin using the blood sample as a reference. Although we identified 8169 SNVs in the tumour (Fig. 3a), the margin appeared genetically normal when compared to blood (Supplementary Fig. 3). About 50% of all SNVs in the tumour were C to T changes (or G to A, opposite strand), most often in NpCpG contexts, corresponding to the mutational signature 1A as described in Alexandrov et al. (Fig. 3b), though we did not observe any kataegis formation. 23 SNVs identified within the coding region of several cancer genes suggest aberrations of the RTK–RAS–PI3K signalling (mutations in PTEN, PIK3R6, and NFI) and MYC signalling pathways (SMARCB1 and LZTR1), and have been recently described in a genomic characterisation of IDH(+) glioma patients. 24 Additionally, complex patterns of somatic structural variants were characteristic of the tumour (Fig. 3c). We observed many translocation and inversion events with additional copy number gains in chromosomes 7, 17, and 20, and copy number losses in chromosomes 1, 6, 9, 10, 11, 18, and 22. Notably, the epigenetic modifiers TET1 (chromosome 10) and DNM1/3B
Links between genomic and epigenomic changes

To investigate links between genetic and epigenetic mutations, we identified SNV sites resulting from the mutation of cytosine to any other base in the tumour, and studied their methylation and hydroxymethylation status in the margin (Fig. 4a). We found that SNV sites were significantly more methylated in the margin compared to non-SNV sites (Mann–Whitney test, \( p \)-value < 2.2e−16, two-sided). Conversely, we found that SNV sites were significantly less hydroxymethylated in the margin compared to non-SNV sites (Mann–Whitney test, \( p \)-value < 2.2e−16, two-sided) with no apparent differences for all types of cytosine mutations (i.e., C to C, C to G, or C to A) (Fig. 4b). Regions identified to be of different ploidy in margin and tumour (Fig. 3c) did not show differences in modification levels (Fig. 4c).

(Chromosomes 19 and 20) were haploid and polyploid in tumour respectively, which suggests a connection between the observed hypermethylation (DNMT polyploidy) and 5hmC loss (diminished TET function) in the tumour.
Tumour purity was estimated to be 71% by genetic analysis and 60% by immunohistochemistry (see Methods). Based on tumour purity, we calculated the expected global level of 5hmC to have a lower limit of 7% (Table 1), assuming complete loss of 5hmC in the tumour. We measured actual levels of 5hmC in the tumour sample to be 1.6%, more than four-fold lower than the estimated lower limit.

DISCUSSION

We used a PCR-free approach to generate accurate methylomes and hydroxymethylomes. Importantly, our ability to resolve 5mC and 5hmC signals reveals global hypermethylation in the tumour, contrary to the appearance of global hypomethylation when using only BS-seq, which actually measures the sum of 5mC and 5hmC and fails to resolve the two signals. Global hypermethylation has just recently also been reported for kidney cancer using an alternative method that was able to distinguish 5mC from 5hmC.25 While genome-wide hypomethylation26, 27 is somewhat regarded as an epigenetic hallmark of tumorigenesis and has been implicated in the evolution of human glioblastomas,28 our results suggest that data obtained from bisulfite-only approaches need to be re-interpreted.

TET2 promoter methylation has previously been observed in low-grade diffuse gliomas lacking IDH1/2 mutations and provides a third mechanism to cause loss in maintenance of 5hmC levels in the tumour. Previous literature has also linked reduced TET function to solid and myeloid malignancies5, 19, 20 and suggested a key role for TET in the prevention of cancer by suppressing cell invasion21 and promoting genome integrity.5, 6 Our results are consistent with this mechanism in the current glioblastoma case.

5mCpG sequences have previously been reported to be mutational hotspots in human genetic disease and cancer-relevant genes, probably through the deamination of 5mC to T.29 Our data show that 5mCpG sites are predisposed to mutations, whereas 5hmCpG sites are protected from mutations during tumorigenesis, mirroring observations recently reported for brain and kidney cancer, and myeloid leukaemias.30 Interestingly, the ‘protection’ at 5hmCpG sites holds for all types of cytosine mutations (C to T, C to G, or C to A).

Our observation that 5hmC levels in the tumour are significantly lower than the estimated lower limit suggests a loss of 5hmC in the genetically "normal" cells within the tumour mass. Earlier studies reported that genome-wide methylation changes in tumours can be acquired in adjacent normal cells linking 5mC to field defects.31–33 The concept of field defects34 has been used to describe the early events in the stepwise transformation of the cancer that can potentially lead to further oncogenic changes. To the best of our knowledge, this is the first observation of acquired, early changes in 5hmC in proximal cells.

Accurate, high resolution whole genome maps have allowed us to discern a role for 5hmC in protecting the genome against somatic mutations, observe global hypermethylation in a tumour and provide new evidence that associates 5hmC with epigenetic transformation in genetically normal cells adjacent to the tumour. Epigenetic analysis of 5hmC, in addition to 5mC, may foster diagnostic approaches in the future.

METHODS

Clinical events
A 71-year old woman underwent supra-total resection of a right frontal intrinsic tumour with maximum diameter of the enhancing component of 19 mm, surrounded by only minimal oedema. Histological examination confirmed the diagnosis of glioblastoma...
multiforme (WHO 9440/3). Tissue collection protocols were compliant with the UK Human Tissue Act 2004 (HTA License ref. 12315) and approved by the Local Regional Ethics Committee (LREC ref. 04/Q0108/60). Informed consent was obtained from the patient before surgery. Surgical samples were taken from the enhancing tumour mass and corresponding non-enhancing margin using advanced surgical techniques described previously. After rapid recovery she underwent six weeks of radiotherapy (60Gy in 30 fractions) with concurrent temozolomide (TMZ) treatment. She then completed two cycles of adjuvant TMZ treatment during 12 weeks. This was stopped because of poor tolerance. While her 9-months post-operative scan was clear of tumour, her 11-months check demonstrated tumour recurrence and she died 3 months later from the combination of tumour progression and thromboembolic complications.

Samples

Brain tumour biopsies were fixed in 10% formalin and embedded in paraffin wax from which 4-µm-thick sections were cut and stained with haemaotoxylin and eosin (H&E). The diagnosis of glioblastoma multiforme (WHO 9440/3) was made by light microscopical examination of H&E-stained sections. Glioblastoma was defined by WHO-2016 criteria of an infiltrative astrocytoma with proliferative activity, necrosis and/or microvascular proliferation.

To determine the IDH1 status by immunohistochemistry, the 4-µm-thick sections were dried at 60 °C for 2 h and further processed on a Bond Max (Leica). After a 60-min pre-treatment with cell conditioner 2 (pH 6) the slides were incubated with 1:80 diluted H09 anti-IDH1 R132H antibody (Dianova, Hamburg, Germany) at

| Sample  | % 5mC | % 5hmC |
|---------|-------|--------|
| Margin  | 52.1  | 17.5   |
| Tumour  | 63.5  | 1.6    |

Table 1. Overall cytosine modification levels of all CpG sites \(n=2.7 \times 10^7\) in a glioblastoma patient
room temperature for 30 min. A standard 3,3’-diaminobenzidine (DAB) detection kit was used for chromogenic detection. The common IDH1 mutation (R132H) was not detected.

To determine the MGMT promoter methylation, H&E-stained slides were reviewed and neoplastic cell-rich tissue was dissected from consecutive unstained sections. DNA was extracted from the dissected tissue using the QiAamp DNA FFPE Tissue Kit (Qiagen) and was bisulphite-converted using the EpiTect Bisulphite Kit (Qiagen). The MGMT promoter methylation was determined by pyrosequencing of four CpG sites (CpGs 76–79) in exon 1 of the MGMT gene using the CE-Marked theerascreen MGMT Pyro Kit on a Pyromark Q24 System (Qiagen). Significant levels of MGMT promoter methylation were not detected.

To determine the tumour cell purity of the samples, H&E-stained slides were reviewed and cell counting was performed on multiple high-powered fields. The tumour cell purity of both the tumour and its margin was assessed by a pathologist prior to molecular profiling. No tumour cells were detected in the margins and the tumour was estimated to consist of 60% tumour cells.

DNA and RNA extraction for sequencing

Each tumour and margin sample was frozen to –80 °C immediately on collection. The frozen samples were then defrosted and a 50 mg portion taken for DNA/RNA extraction. The samples were homogenised and then divided for DNA and RNA extraction. DNA extraction of both tissues and blood was performed using the QIAGEN DNeasy Blood and Tissue Kit. RNA extraction was performed using the QIAGEN RNeasy Mini Kit. Quantification was performed using the Qubit quantification assay. The tumour/normal pair used in this study was the sample judged most pure by comparison with blood DNA using whole genome sequencing.

PCR-free oxBS and BS library preparation and sequencing

PCR-free ReBuilT libraries for bisulphite sequencing were generated following our previously published method. For the generation of PCR-free libraries for oxBS, libraries were oxidised (TrueMethyl Kit from CEGX) prior to bisulphite conversion using the EpiTect Bisulphite Kit (Qiagen). The oxBS library preparation and sequencing were generated following our previously published method. The generation of PCR-free libraries for oxBS, libraries were oxidised (TrueMethyl Kit from CEGX) prior to bisulphite treatment. In brief, 350 and 175 ng of sonicated DNA from tumour and margin respectively were end repaired and dA-tailed before ligation of customized adapter pair. After (oxidation)-bisulphite treatment, degraded fragments were recovered through primer extension before dA-tailing and second adapter ligation. Libraries were subsequently immobilized on streptavidin-coated magnetic beads, washed with binding buffer, eluted with 50 mM NaOH at 60 °C for 15 min. BS and oxBS-seq libraries were sequenced on an illumina HiSeq 2500 platform, V4 chemistry, 2×101+7 index cycles.

RNA-seq library preparation and sequencing

Libraries of the tumour and margin samples were prepared using TruSeq RNA Access kit (Illumina), using an input of 40 and 20 ng respectively. The libraries were sequenced for 2×75 cycles on HiSeq 2500 in Rapid mode achieving an average coding coverage of 108x (F2), 136x (F3), 277x (M1), 272x (M3).

Computational analysis

Code availability. Details of the analysis are available in the manuscript’s GitHub repository. https://github.com/sblab-bioinformatics/epigenetics-of-glioblastoma.

DNA-seq data analysis. Libraries were sequenced on an Illumina HiSeq 2500 platform. Alignment to the human reference genome GRCh37 and quality control was performed using Isaac. Identification of somatic SNVs and small somatic indels (<50 bp) was performed by Strelka. Large copy number variants and structural variants were respectively called with Canvas and Manta. The full workflow can be found in the Tumour Normal application of the Illumina BaseSpace platform. http://support.illumina.com/content/dam/illumina-support/help/BaseSpace_App_TumorNormal_v2/tumor-normal-v2-help.html.

BS and oxBS-seq data analysis. Raw reads were trimmed to remove adapter sequences ligated to the 3’-end using cutadapt and aligned to the human reference genome hg19 using bwa-mem, a wrapper around bwa-mem.

RNA-seq data analysis. As above, reads were trimmed to remove adapter sequences using cutadapt and aligned to reference transcripts obtained from Ensembl and quantified using kallisto.

Tumour purity estimation. The copy number variant analysis using Canvas provided us with an estimate of 71% purity of the tumour sample. This software calculates coverage and tumour SNV allele frequencies at heterogeneous germine positions along the genome. It then assigns copy numbers per genomic regions and infers genome-wide ploidy and purity by fitting the data to expected models for each copy number state given purity and diploid coverage level combinations. Purity is then derived from the best fitting model.

Data availability

Data are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-5171

Acknowledgements

The Balasubramanian laboratory is supported by core funding from Cancer Research UK (C14303/A17197). S.B. is a Senior Investigator of the Wellcome Trust (grant no. 099232/z/12/z). Colin Watts is supported by funding from CRUK and The Brain Tumour Charity.

Author contributions

All authors contributed to the conception and design of the experiments. T. S. and K. A. provided samples and clinical data. E.A.R., G.R.M., S.F., T.J., M.L. and Z.K. performed the experiments. D.B., S.M.C. and J.B. performed computational analyses. All authors analysed and interpreted the data. E.A.R., D.B., S.M.C. and S.B. wrote the manuscript, with contributions from all authors.

Competing interests

S.B. is a founder and shareholder of Cambridge Epigenetics Ltd. The authors otherwise declare no competing interests.

References

1. Baylin, S. B. et al. Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. Hum. Mol. Genet. 10, 687–692 (2001).
2. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324, 929–930 (2009).
3. Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324, 930–935 (2009).
4. Pfeifer, G. P., Kadam, S. & Jin, S.-G. 5-Hydroxymethylcytosine and its potential roles in development and cancer. Oncogene 32, 687–692 (2011).
10. Kudo et al. Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation. Cancer Sci. 103, 670–676 (2012).
11. Scozzic, L., Mouly, E. & Bernard, O. A. TET proteins and the control of cytosine demethylation in cancer. Genome Med. 7, 9 (2015).
12. McInroy, G. R. et al. Enhanced methylation analysis by recovery of unsequenceable fragments. PLoS One 11, e0152322 (2016).
13. Booth, M. J. et al. Quantitative sequencing of 5-methylcytosine and 5 hydroxymethylcytosine at single-base resolution. Science 336, 934–937 (2012).
14. Mullighan, C. G. TET2 mutations in myelodysplasia and myeloid malignancies. Clin. Cancer Res. 18, 5562–5571 (2012).
15. Xu, W. Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. Oncogene 32, 663–669 (2012).
16. Delhommeau, F. et al. Mutation in TET2 in myeloid cancers. N. Engl. J. Med. 360, 2289–2301 (2009).
17. Mullighan, C. G. TET2 mutations in myelodysplasia and myeloid malignancies. Nat. Genet. 41, 766–767 (2009).
18. Yang, H., Ye, D., Guan, K. L. & Xiong, Y. IDH1 and IDH2 mutations in tumorigenesis: mechanistic insights and clinical perspectives. Clin. Cancer Res. 18, 5562–5571 (2012).
19. Yang, H. et al. Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. Oncogene 32, 663–669 (2012).
20. Huang, Y. & Rao, A. Connections between TET proteins and aberrant DNA modification in cancer. Trends Genet. 30, 1–11 (2014).
21. Huang, Y. & Rao, A. Connections between TET proteins and aberrant DNA modification in cancer. Trends Genet. 30, 1–11 (2014).
22. Hsu, C. H. et al. TET1 suppresses cancer invasion by activating the tissue inhibitors of Metalloproteinases. Cell Rep. 2, 568–579 (2012).
23. Pfeifer, G. P. Mutagenesis at methylated CpG sequences. Curr. Top. Microbiol. Immunol. 301, 259–281 (2006).
24. Mazor, T. et al. DNA methylation and somatic mutations converge on the cell cycle and define similar evolutionary histories in brain tumors. Cancer Cell 28, 307–317 (2015).
25. Mazor, T. et al. DNA methylation and somatic mutations converge on the cell cycle and define similar evolutionary histories in brain tumors. Cancer Cell 28, 307–317 (2015).
26. Pfeifer, G. P. Mutagenesis at methylated CpG sequences. Curr. Top. Microbiol. Immunol. 301, 259–281 (2006).
27. Ehrlich, M. DNA methylation in cancer: too much, but also too little. Oncogene 21, 5400–5413 (2002).
28. Mazor, T. et al. DNA methylation and somatic mutations converge on the cell cycle and define similar evolutionary histories in brain tumors. Cancer Cell 28, 307–317 (2015).
29. Kriaucionis, S. & Schuster-Boeckler, B. 5-hydroxymethylcytosine maps regions with reduced mutation frequency in human DNA. Elife 5, 459–466 (2016).
30. Tomkova, M., McClellan, M., Kriaucionis, S. & Schuster-Boeckler, B. 5-hydroxymethylcytosine marks regions with reduced mutation frequency in human DNA. Elife 5, 459–466 (2016).
31. Suter, C. M., Martin, D. I. & Ward, R. I. Hypomethylation of L1 retrotransposons in colorectal cancer and adjacent normal tissue. Int. J. Colorectal Dis. 19, 95–101 (2004).
32. Teschendorff, A. E. et al. DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. Nat. Commun. 7, 10478 (2016).
33. Katsurano, M. et al. Early-stage formation of an epigenetic field defect in a mouse colitis model, and non-essential roles of T- and B-cells in DNA methylation induction. Oncogene 31, 342–351 (2012).
34. Slaughter, D. P., Southwick, H. W. & Smedkaj, W. Field cancerization in oral stratified squamous epithelium: clinical implications of multicentric origin. Cancer 6, 963–968 (1953).
35. Piccirillo, S. G. M. et al. Fluorescence-guided surgical sampling of glioblastoma identifies phenotypically distinct tumour-initiating cell populations in the tumour mass and margin. Br. J. Cancer 107, 462–468 (2012).
36. Racy, C. et al. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. Bioinformatics 29, 2041–2043 (2013).
37. Saunders, C. T. et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 28, 1811–1817 (2012).
38. Roller, E., Ivakhno, S., Lee, S., Royce, T. & Tanner, S. Canvas: versatile and scalable detection of copy number variants. Bioinformatics 32, 2375–2377 (2016).
39. Chen, X. et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 32, 1220–1222 (2016).
40. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17, 10 (2011).
41. Pedersen, B. S., Eyring, K., De, S., Yang, I. V. & Schwartz, D. A. Fast and accurate alignment of long bisulfite-sequ reads. arXiv 1401.1129, 1–2 (2014).
42. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303.3997, 1–3 (2013).
43. R reality, C. M. et al. Ensembl 2016. Nucleic Acids Res. 44, D710–D716 (2016).
44. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34, 525–527 (2016).
45. Forbes, S. A. et al. COSMIC: exploring the world’s knowledge of somatic mutations in human cancer. Nucleic Acids Res. 43, D805–D811 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/