



## TECHNICAL NOTE

# Ewastools: Infinium Human Methylation BeadChip pipeline for population epigenetics integrated into Galaxy

Katarzyna Murat<sup>1</sup>, Björn Grüning <sup>2</sup>, Paulina Wiktoria Poterłowicz<sup>3</sup>, Gillian Westgate<sup>1</sup>, Desmond J. Tobin<sup>4</sup> and Krzysztof Poterłowicz <sup>1,\*</sup>

<sup>1</sup>Center for Skin Sciences, University of Bradford, Richmond Road, Bradford BD7 1DP, UK; <sup>2</sup>Freiburg Galaxy Team, University of Freiburg, Fahrenbergplatz, 79085 Freiburg im Breisgau, Germany; <sup>3</sup>Hull York Medical School, University of York, University Rd, York YO10 5DD, UK and <sup>4</sup>The Charles Institute for Dermatology, Belfield, School of Medicine, University College Dublin, Ireland

\*Correspondence address. Krzysztof Poterłowicz, Center for Skin Sciences, University of Bradford, Richmond Road, Bradford BD7 1DP, UK. E-mail:

k.poterlowicz1@bradford.ac.uk  <http://orcid.org/0000-0001-6173-5674>

## Abstract

**Background:** Infinium Human Methylation BeadChip is an array platform for complex evaluation of DNA methylation at an individual CpG locus in the human genome based on Illumina's bead technology and is one of the most common techniques used in epigenome-wide association studies. Finding associations between epigenetic variation and phenotype is a significant challenge in biomedical research. The newest version, HumanMethylationEPIC, quantifies the DNA methylation level of 850,000 CpG sites, while the previous versions, HumanMethylation450 and HumanMethylation27, measured >450,000 and 27,000 loci, respectively. Although a number of bioinformatics tools have been developed to analyse this assay, they require some programming skills and experience in order to be usable. **Results:** We have developed a pipeline for the Galaxy platform for those without experience aimed at DNA methylation analysis using the Infinium Human Methylation BeadChip. Our tool is integrated into Galaxy (<http://galaxyproject.org>), a web-based platform. This allows users to analyse data from the Infinium Human Methylation BeadChip in the easiest possible way. **Conclusions:** The pipeline provides a group of integrated analytical methods wrapped into an easy-to-use interface. Our tool is available from the Galaxy ToolShed, GitHub repository, and also as a Docker image. The aim of this project is to make Infinium Human Methylation BeadChip analysis more flexible and accessible to everyone.

**Keywords:** Infinium Human Methylation BeadChip; epigenome-wide association studies (EWAS); DNA methylation; Galaxy Project; pipeline; sequence analysis

## Background

Over the past several years comprehensive sequencing datasets have been generated, allowing analysis of genome-wide activity in cohorts of different individuals to be increasingly available. Infinium Human Methylation BeadChip requires only a few days to produce methylome profiles of human samples with a low

sample input requirement (as low as 500 ng of genomic DNA) for the starting material [1]. Studies performed recently have identified variation naturally occurring in the genome associated with disease risk and prognosis, including tumour pathogenesis [2]. This raised interest in the concept of epigenome-wide association studies (EWAS). The term “epigenome” means “on top of the

**Received:** 14 March 2019; **Revised:** 24 February 2020; **Accepted:** 21 April 2020

© The Author(s) 2020. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

genome” and refers to specific changes in genome regulatory activity occurring in response to environmental stimuli [3]. Epigenetic modifications do not change the underlying DNA sequence but can cause multiple changes in gene expression and cellular function [4]. In humans, DNA methylation occurs by attaching a methyl group to the cytosine residue. This has been suggested as a suppressor of gene expression [5]. Multiple methods for DNA methylation analysis were developed, including PCR and pyrosequencing of bisulfite converted DNA, dedicated to studying a small number of methylation sites across a number of samples [6]. Assays like whole-genome bisulfite sequencing and reduced representation bisulfite sequencing allow global quantification of DNA methylation levels. However, running this type of analysis for a larger number of samples can be prohibitively laborious and expensive [7]. The Infinium Human Methylation BeadChip [1] offers unprecedented applicability and affordability owing to the low costs of reagents, short time of processing, high accuracy, and low input DNA requirements. It determines quantitative array-based methylation measurements at the single-CpG-site level of >850,000 loci [8], covering most of the promoters and also numerous other loci. This makes this assay suitable for systematic investigation of methylation changes in normal and diseased cells [3]. As such it has become one of the most comprehensive solutions on the market [9]. However, Illumina commercial software generates additional costs and is not suitable for everyone. Therefore there is a need to create freely available software able to perform comprehensive analysis including quality control, normalization, and detection of differentially methylated regions (DMRs) [9]. Open source software packages (e.g., DMRcate [10], Minfi [11], ChAMP [12], methylumi [13], RnBeads [14]) require high-performance computational hardware as well as command line experience in order to run the analysis. This is why one of the aims of the our Infinium Human Methylation BeadChip pipeline was to set and implement these methods into a user-friendly environment. The tool has been developed to provide users with an enhanced understanding of the Infinium Human Methylation BeadChip analysis. The workflow includes methods for preprocessing with a stratified quantile normalization preprocess. It includes quantile or extended implementation of the functional normalization preprocess Funnorm with unwanted variation removal, sample-specific quality assessment, and methodology for calling DMR and differentially methylated position (DMP) detection. Scripts were combined and published on the web-based platform Galaxy, a graphical interface with tools and ready-to-run workflows providing a solution for non-programmer scientists to analyse their data and share their experience with others [15]. Configuration files are publicly shared on our GitHub repository [16], with code and dependency settings also available to download and install via the Galaxy ToolShed [17]. Our tool was created and tested using Planemo [18], an integrated workspace for Galaxy tool development with a default configuration and shed tool set-up available via Docker (operating system-level virtualization) [16].

## Tool Description

The workflow combines 5 main steps (see Fig. 1), starting with raw intensity data loading (`.idat`) and then preprocessing and optional normalization of the data. The next quality control step performs an additional sample check to remove low-quality data, which normalization cannot detect. The workflow gives the user the opportunity to perform any of these preparation and data-cleaning steps, including the next highly recommended

genetic variation annotation step resulting in single-nucleotide polymorphism (SNP) identification and removal. Finally, the dataset generated through all of these steps can be used to find DMPs and DMRs with respect to a phenotype covariate. All the steps, as well as simple preparation and analysis options, are shown in Fig. 2 and explained in detail below.

## Data loading

The Infinium Human Methylation BeadChip assay interrogates fluorescent signals (green and red) from the methylated and unmethylated sites into binary values that can be read directly as IDAT files [1]. Illumina’s GenomeStudio (GenomeStudio, RRID:SC R\_010973) solution converts the data into plain-text ASCII files, losing a large amount of information during this process [19]. To prevent this kind of data loss we introduced an R-based .IDAT file-loading method, which is a combination of `illuminaio readIDAT` and `minfi RGChannelSet` functions. It reads intensity information from both treatment and control data and based on this it builds up a specific joined dataset.

## Preprocessing and normalization

Green and red channel signals from .IDAT files can be converted into methylated and unmethylated signals assigned to methylation levels or  $\beta$  values.  $\beta$  are built in `RatioSet` object, and they estimate the methylation level using channel ratios in a range between 0 and 1, with 0 being unmethylated and 1 being fully methylated [19]. However, these 2 classes can also be preprocessed and normalized with 2 methods available [19]. `PreprocessQuantile` implements stratified quantile normalization preprocessing and is supported for small changes such as in 1-type samples, e.g., blood datasets. In contrast, `preprocessFunnorm` is aimed at global biological differences such as healthy and occurred datasets with different tissue and cell types. This is called the “between-array normalization method” and removes unwanted variation [19]. In addition unwanted probes containing either an SNP at the CpG interrogation or at the single-nucleotide extension can be removed (recommended) [19].

## Quality assessment and control

Data quality assurance is an important step in Infinium Human Methylation BeadChip analysis. The quality control function extracts and plots the data frame with 2 columns `mMed` and `uMed`, which are the medians of `MethylSet` signals (`Meth` and `Unmeth`). Comparing these against one another allows users to detect and remove low-quality samples that normalization cannot correct [11].

## Annotating probes affected by genetic variation

SNP regions may affect results of downstream analysis. The `RemoveSNPs` step returns data frames containing the SNP information of unwanted probes and removes them from the dataset [19].

## DMP and DMR identification

The main goal of the Infinium Human Methylation BeadChip tool is to simplify the way differentially methylated locus sites are detected. The workflow contains a function detecting DMPs with respect to the phenotype covariate, and a method

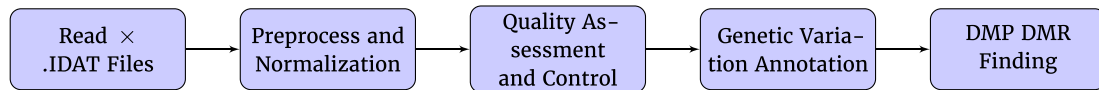


Figure 1: Simplified workflow for analysing epigenetics data.

**Infinium Human Methylation BeadChip** Determines differentially methylated regions and positions from idat files (Galaxy Version 2.1.0) ☆ Favorite ▼ Options

**Red .IDAT files**

- 8: GSM1588707\_8795207119\_R06C02\_Red.idat
- 7: GSM1588707\_8795207119\_R06C02\_Grn.idat
- 6: GSM1588706\_8795207135\_R02C02\_Red.idat
- 5: GSM1588706\_8795207135\_R02C02\_Grn.idat
- 4: GSM1588705\_8795207119\_R05C02\_Red.idat
- 3: GSM1588705\_8795207119\_R05C02\_Grn.idat
- 2: GSM1588704\_8795207135\_R01C02\_Red.idat

Red .IDAT files extension is followed by the unmethylated signal intensity read in the red channel.

**Green .IDAT files**

- 8: GSM1588707\_8795207119\_R06C02\_Red.idat
- 7: GSM1588707\_8795207119\_R06C02\_Grn.idat
- 6: GSM1588706\_8795207135\_R02C02\_Red.idat
- 5: GSM1588706\_8795207135\_R02C02\_Grn.idat
- 4: GSM1588705\_8795207119\_R05C02\_Red.idat
- 3: GSM1588705\_8795207119\_R05C02\_Grn.idat
- 2: GSM1588704\_8795207135\_R01C02\_Red.idat

Green .IDAT files extension is followed by the methylated signal intensity read in the green channel.

**(Optional) Preprocessing Method**

No Selection (use default)

Mapping Illumina methylation array data to the genome with or without additional preprocess.

**Phenotype Table**

10: ucsc.gtf

Table of compared probes and their characteristics, may be categorical (e.g. cancer vs. normal) or continuous (e.g. blood pressure).

**maxGap Size**

250

If cluster is not provided this maximum location gap will be used to define cluster.

**Cutoff Size**

0.1

A numeric value. Values of the estimate of the genomic profile above the cutoff or below the negative of the cutoff will be used as candidate regions. It is possible to give two separate values (upper and lower bounds). If one value is given, the lower bound is minus the value.

**Number of Resamples**

0

An integer denoting the number of resamples to use when computing null distributions. This defaults to 0. If permutations is supplied that defines the number of permutations/bootstraps and B is ignored.

**null Method**

permutation

Method used to generate null candidate regions (defaults to 'permutation'). Note that for cases with more than one covariate the permutation approach is not generally recommended.

**Phenotype Type**

categorical

Figure 2: Screenshot from the Galaxy interface, showing Infinium Human Methylation BeadChip workflow as discussed in the case study section.

for finding DMRs [11]. DMRs can be tracked using a bump-hunting algorithm. The algorithm first implements a t-statistic at each methylated locus location, with optional smoothing, then groups probes into clusters with a maximum location gap and a cut-off size to refer the lowest possible value of genomic profile hunted by our tool [20].

### Functional annotation and visualization

In addition to downstream analysis, users can access annotations provided via Illumina by ChIPpeakAnno (ChIPpeakAnno, [RRID:SCR.012828](#)) annoPeaks tool [19] or perform additional functional annotations using the Gene Ontology (GO) via Cluster Profiler GO tool. The GO tool provides a very detailed representation of functional relationships between biological processes, molecular function, and cellular components across data [21]. Once specific regions have been chosen, Cluster Profiler GO visualizes enrichment results (see Fig. 3). Many researchers use annotation analysis to characterize the function of genes, which highlights the potential for Galaxy to be a solution for wide-ranging multi-omics research.

### Documentation and training

We have also provided training sessions and interactive tours for user self-learning. The training materials are freely accessible at the Galaxy (Galaxy, [RRID:SCR.006281](#)) project Github repository [22]. Such training and tours guide users through an entire analysis. The following steps and notes help users to explore and better understand the concept. Slides and hands-on instruction describe the analysis workflow, and all necessary input files are ready to use via Zenodo [23], as well as a Galaxy Interactive Tour, and a tailor-made Galaxy Docker image for the corresponding data analysis.

### Case study

Compared to genetic studies EWAS provides a unique opportunity to study dynamic response to treatment. It has been suggested that DNA methylation is associated with drug resistance [24]. To validate our suite we have performed analysis of differentially methylated regions using publicly available data from the Infinium Human Methylation BeadChip array of melanoma biopsies before and after mitogen-activated protein

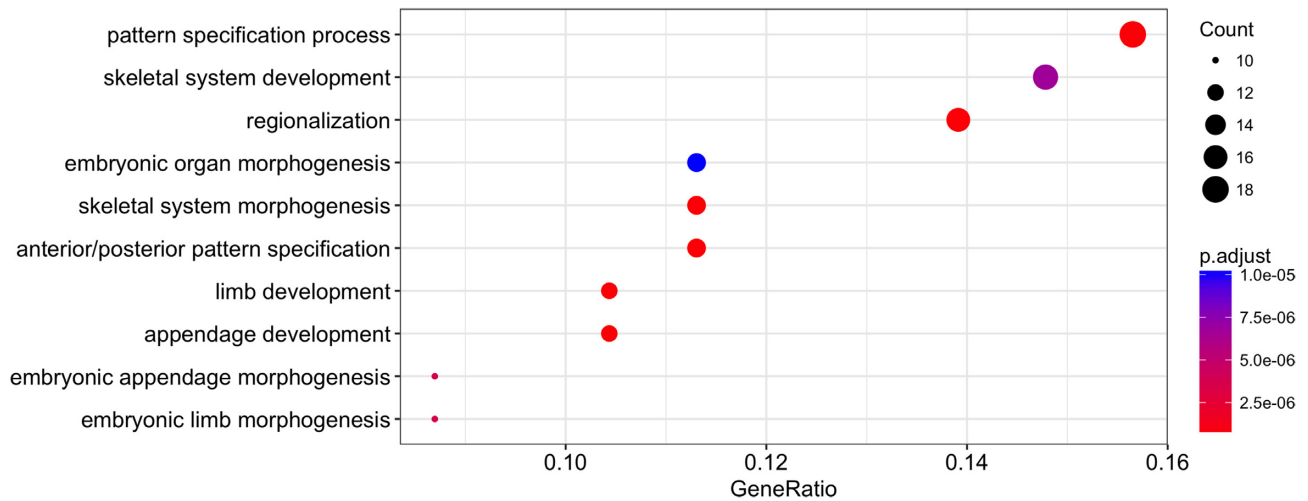


Figure 3: Functional annotation of DMRs found in melanoma biopsies before and after MAPKi treatment.

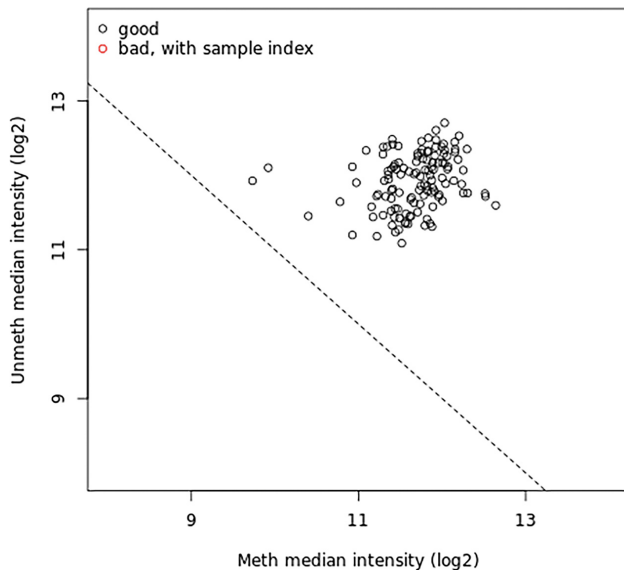


Figure 4: Quality control plot representing median intensity of melanoma in samples before and after MAPKi treatment. The plot compares median total intensity ( $\log_2$ ) of the methylated channel (x-axis) and unmethylated channel (y-axis). Bad-quality samples fall under the threshold and are colored red. There is no bad-quality sample in this study

kinase inhibitor (MAPKi) treatment [25], obtained from the Gene Expression Omnibus (GEO) (GSE65183). Methylation profiling by genome tiling array in melanoma can help us understand how non-genomic and immune changes can have an impact on treatment efficiency and disease progression. Raw image IDAT files were loaded into the Galaxy environment using Data Libraries. EWAS workflow was run on Red and Green dataset collections of patient-matched melanoma tumours biopsied before therapy and during disease progression. The IDAT files, predefined phenotype tables, and up-to-date genome tables (UCSC Main on Human hg19 Methyl450) [16] were used as inputs. To detect poorly performing samples we ran quality diagnostics. The provided samples passed the quality control test (in Fig. 4) because they clustered together with higher median intensities, confirming their good quality [19]. Differentially methylated loci were identified using single-probe analysis implemented by our tool

with the following parameters: phenotype set as “categorical” and qCutoff size set to 0.05. The bump-hunting algorithm was applied to identify DMRs with maximum location gap parameter set to 250, genomic profile above the cut-off equal to 0.1, number of resamples set to 0, and null method set to “permutation and verbose equal FALSE,” which means that no additional progress information will be printed. Differentially Methylated Regions and Positions revealed the need for further investigation of tissue diversity in response to environmental changes [26]. The nearest transcription start sites found in the gene set can be listed as follows: PITX1, SFRP2, MSX1, MIR21, AXIN2, GREM1, WT1, CBX2, HCK, GTSE1, SNCG, PDPN, PDGFRA, NAF1, FGF5, FOXE1, THBS1, DLK1, and HOX gene family. The results of the re-analysis are available in the GitHub repository [27].

### Important findings

Although hypermethylated genes identified by “EWAS-suite” have been previously associated with cancer, this is the first time a link between them and MAPKi treatment resistance is reported. This data demonstrates the presence of platelet-derived growth factor receptor (PDGFR), which is suggested to be responsible for RAS/MAPK pathway signaling. Trough activation may regulate the MAPKi mechanism in non-responsive tumours. The methylation regulation of this altered status of PDGFR requires additional studies [25]. The PITX1 suppressor gene was found as one of the factors decreasing gene expression in human cutaneous malignant melanoma and might contribute to progression and resistance via promoting cell proliferative activity [28]. It has been found that the homeodomain transcription factor MSX1 and the CBX2 polycomb group protein are likely to be treatment resistance factors and are reported as downregulated and inactivated in melanoma tumours [29]. Previous published studies are limited to local surveys and serial biopsies. Thus, the stimulus of innate or acquired MAPKi resistance may be linked to epigenetics. GO annotation provides information regarding the function of genes [30]. GO analysis identified the pattern specification process (GO:0007389), skeletal system development (GO:0001501), and regionalization (GO:000300) as significantly overrepresented categories within the above DMRs, suggesting that melanoma MAPKi resistance could be related to the cells’ developmental process within specific environments.

## Conclusion

With the rapidly increasing volume of epigenetics data available, computer-based analysis of heritable changes in gene expression becomes more and more feasible. Many genome-wide epigenetics studies have focused on generation of data, with data interpretation now being the challenge. Risk evaluation, disease management, and novel therapeutic development are prompting researchers to find new bioinformatic frameworks and approaches. In this regard we propose a user-friendly tool suite available via Galaxy platform. Ewastools allows life scientists to run complex epigenetics analysis [16]. The case study presented provides a tangible example of how population epigenetics analysis can provide additional insights into melanoma therapeutic resistance.

## Availability of Source Code and Requirements

- Project name: Ewastools: Infinium Human Methylation Bead-Chip pipeline for population epigenetics integrated into Galaxy
- Project home page: [https://github.com/kpbioteam/ewas\\_galaxy](https://github.com/kpbioteam/ewas_galaxy)
- Operating system(s): Linux (recommended), Mac
- Programming language: R programming language (version 3.3.2, x86 64bit)
- Other requirements: Galaxy [22], Docker [31]
- License: MIT License
- biotoolsID identifier: <https://bio.tools/ewastools>
- RRID:SCR.018085

## Availability of Supporting Data and Materials

The test dataset in this article is available in the GEO database under accession GSE65186. The results of the re-analysis of the GSE65186 dataset are available in the GitHub repository ([https://github.com/kpbioteam/ewastools-case\\_study](https://github.com/kpbioteam/ewastools-case_study)). All tools described here are available in the Galaxy ToolShed (<https://toolshed.g2.bx.psu.edu>). The Dockerfile required to automatically deploy the pre-built Docker image is available at <https://galaxyproject.org/use/ewas-galaxy/>. Archival snapshots of the code are available in the GigaScience GigaDB repository [32].

## Abbreviations

DMP: differentially methylated position; DMR: differentially methylated region; EWAS: epigenome-wide association study; GEO: Gene Expression Omnibus; GO: Gene Ontology; MAPKi: mitogen-activated protein kinase inhibitor; PDGFR: platelet-derived growth factor receptor; SNP: single-nucleotide polymorphism; UCSC: University of California Santa Cruz.

## Competing Interests

The authors declare that they have no competing interests.

## Authors' Contributions

K.P. conceived and designed the study, K.P., K.M. and B.G. developed the software, K.P., K.M., P.W.P. and B.G. did testing, K.P., K.M. and P.W.P. performed the analyses, K.P., K.M., D.J.T. and G.W. provided biological interpretation. All authors wrote the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

We thank Michal Gdula for constructive criticism of the manuscript.

## References

1. Infinium Methylation Assay Overview. 2020. <https://emea.illumina.com/science/technology/microarray/infinium-methylation-assay.html>. Accessed on 15 February 2020.
2. Lee JJ, Murphy GF, Lian CG. Melanoma epigenetics: novel mechanisms, markers, and medicines. *Lab Invest* 2014;**94**(8):822–38.
3. Rakyan VK, Down TA, Balding DJ, et al. Epigenome-wide association studies for common human diseases. *Nat Rev Genet* 2011;**12**(8):529–41.
4. Egger G, Liang G, Aparicio A, et al. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 2004;**429**(6990):457–63.
5. Klose RJ, Bird AP. Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* 2006;**31**(2):89–97.
6. Sandoval J, Heyn H, Moran S, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* 2011;**6**(6):692–702.
7. Kristensen LS, Hansen LL. PCR based methods for detecting single locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment. *Clin Chem* 2009;**55**(8):1471–83.
8. Pidsley R, Wong CCY, Volta M, et al. A data driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* 2013;**14**(1):293.
9. Marabita F, Almgren M, Lindholm ME, et al. An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform. *Epigenetics* 2013;**8**(3):333–46.
10. Peters TJ, Buckley M, Statham AL, et al. DMRcate Illumina 450 K methylation array apatial analysis methods. 2014. <https://bioconductor.org/packages/release/bioc/html/DMRcate.html>. Accessed 20 November 2018.
11. Hansen KD, Aryee M. minfi: Analyze Illumina's 450k methylation arrays. 2012. <https://bioconductor.org/packages/release/bioc/html/minfi.html>.
12. Morris TJ, Butcher LM, Feber A, et al. ChAMP 450k chip analysis methylation pipeline. *Bioinformatics* 2013;**30**(3):428–30.
13. Davis S, Du P, Bilke S, et al. methylumi: Handle Illumina methylation data.. 2019. [lease/bioc/html/methylumi.html](https://bioconductor.org/packages/release/bioc/html/methylumi.html). Accessed on 20 January 2019.
14. Assenov Y, Muller F, Lutsik P, et al. Comprehensive analysis of DNA methylation data with RnBeads. *Nat Methods* 2014;**11**(11):1138–548.
15. Goecks J, Nekrutenko A, Taylor J, Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 2010;**11**(8), doi:10.1186/gb-2010-11-8-r86.
16. Murat K, Poterlowicz K. Source Code of EWAS Tools. 2018. <https://github.com/kpbioteam>. Accessed on 20 February 2020.
17. Murat K, Poterlowicz K. Published Tools. 2018. [https://toolshed.g2.bx.psu.edu/repository?repository\\_id=f706255627649ca9](https://toolshed.g2.bx.psu.edu/repository?repository_id=f706255627649ca9). Accessed 20 February 2020.
18. Planemo documentation. 2019. <https://planemo.readthedocs.io/en/latest/>. Accessed 20 February 2020.

19. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014;**30**(10):1363–9.
20. Jaffe AE, Murakami P, Lee H, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol* 2012;**41**(1):200–9.
21. Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 2004;**32**(suppl.1):D258–61.
22. Murat K, Poterlowicz K. EWAS suite training. 2018. <https://galaxyproject.github.io/training-material/topics/epigenetics/tutorials/ewas-suite/tutorial.html>, Accessed 20 February 2020.
23. Murat K, Poterlowicz K. Training data for ‘ewas\_suite’ analysis. Zenodo 2018, doi:10.5281/zenodo.1251211.
24. Verma M. Genome-wide association studies and epigenome-wide association studies go together in cancer control. *Future Oncol* 2016;**12**(13):1645–64.
25. Hugo W, Shi H, Sun L, et al. Non genomic and immune evolution of melanoma acquiring MAPKi resistance. *Cell* 2015;**162**(6):1271–85.
26. Bock C, Lengauer T. Computational epigenetics. *Bioinformatics* 2008;**24**(1):1–10.
27. EWAStools case study repository. <https://github.com/kpbioteam/ewastools-case.study>. Accessed 15 February 2020
28. Osaki M, Chinen H, Yoshida Y, et al. Decreased PITX1 gene expression in human cutaneous malignant melanoma and its clinicopathological significance. *Eur J Dermatol* 2013;**23**(3):344–9.
29. Clermont PL, Sun L, Crea F, et al. Genotranscriptomic meta analysis of the Polycomb gene CBX2 in human cancers initial evidence of an oncogenic role. *Br J Cancer* 2014;**111**(8):1663–532.
30. Ashburner M, Ball CA, Blake JA, et al. Gene Ontology tool for the unification of biology. *Nat Genet* 2000;**25**(1):25.
31. Docker documentation. 2017. <https://media.readthedocs.org/pdf/docker-sean/latest/docker-sean.pdf>. Accessed 20 February 2020.
32. Murat K, Grüning B, Poterlowicz PW, et al. Supporting data for “Ewastools: Infinium Human Methylation BeadChip pipeline for population epigenetics integrated into Galaxy.” Giga-Science Database 2020. <http://dx.doi.org/10.5524/100744>.
33. Peters, TJ., Buckley, MJ., Statham, AL., Pidsley, R., Samaras, K., V Lord, R., Clark, SJ. and Molloy, PL. (2015) De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin* **8**, 6 10.1186/1756-8935-8-625972926