eFORGE: a tool for identifying cell type-specific signal in epigenomic data

Charles E. Breeze[1#], Dirk S. Paul[1], Jenny van Dongen[2], Lee M. Butcher[1,3], John C. Ambrose[1], James E. Barrett[1], Robert Lowe[4], Vardhman K. Rakyan[4], Valentina Iotchkova[5], Mattia Frontini[6,7,8], Kate Downes[6,7], Willem H. Ouwehand[6,7,8,9], Anke Bergmann[10,11], Reiner Siebert[10], Edo Vellenga[12], Sadia Saeed[13,14], Filomena Matarese[14], Joost H.A. Martens[14], Hendrik G. Stunnenberg[14], Javier Herrero[1], Ewan Birney[5], Ian Dunham[5], and Stephan Beck[1#]

[1]UCL Cancer Institute, University College London, London, WC1E 6BT, UK
[2]Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[3]Department of Surgery and Cancer, Imperial College London, London, W12 0NN, UK
[4]Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, E1 2AT, London, UK
[5]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK
[6]Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Long Road Cambridge, CB2 0PT, UK
[7]National Health Service (NHS) Blood and Transplant, University of Cambridge, Cambridge Biomedical Campus, Long Road, Cambridge, CB2 0PT, UK
[8]British Heart Foundation Centre of Excellence, Cambridge Biomedical Campus, Long Road, Cambridge, CB2 0QQ, UK
[9]Department of Human Genetics, The Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1HH, UK
[10]Institute of Human Genetics, Christian Albrechts University, Kiel, Germany
[11]Department of Pediatrics, Christian-Albrechts-University Kiel & University Hospital Schleswig-Holstein, Kiel, Germany
[12]Department of Hematology, University of Groningen and University Medical Center Groningen, P.O. Box 30001, 9700 RB Groningen, the Netherlands
[13]Department of Biochemistry, PMAS Arid Agriculture University Rawalpindi, 46300 Rawalpindi, Pakistan
[14]Radboud University, Department of Molecular Biology, Faculty of Science, Nijmegen Centre for Molecular Life Sciences, 6500 HB Nijmegen, the Netherlands
 # Correspondence to CEB (c.breeze@ucl.ac.uk) and SB (s.beck@ucl.ac.uk)

Abstract
Epigenome-wide association studies (EWAS) provide an alternative approach for studying human disease through consideration of non-genetic variants such as altered DNA methylation. However, analysis of EWAS data remains challenging. We therefore developed eFORGE (http://eforge.cs.ucl.ac.uk/), a new standalone and web-based tool for the analysis and interpretation of such data. eFORGE determines the cell type-specific regulatory component of a set of EWAS-identified differentially methylated positions. This is achieved by detecting enrichment of overlap with DNase I hypersensitive sites across 454 samples (tissues, primary cell types and cell lines) from the ENCODE, Roadmap Epigenomics and BLUEPRINT projects. Application of eFORGE to 18 publicly available EWAS datasets identified disease-relevant cell types for several common diseases, a stem cell-like signature in cancer, and demonstrated the ability to detect cell composition effects for EWAS performed on heterogeneous samples. Our approach bridges the gap between data from current large-scale epigenomics projects and EWAS-derived target selection to yield insight into disease aetiology.