# Supplementary Note

# GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals

Valentina Iotchkova[1,2], Graham R.S. Ritchie[1,2], Matthias Geihs[1], Sandro Morganella[2], Josine L. Min[3], Klaudia Walter[1], Nicholas Timpson[3], UK10K Consortium[4], Ian Dunham[2], Ewan Birney[2,*] and Nicole Soranzo[1,5,6,*]

1. Human Genetics, Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1HH, United Kingdom; 2. European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom; 3. MRC Integrative Epidemiology Unit, Avon Longitudinal Study of Parents and Children, Population Health Science, Bristol Medical School, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, United Kingdom; 4. A full list of authors can be found in the Supplementary Note; 5. Department of Haematology, University of Cambridge, Cambridge CB2 0AH, United Kingdom; 6. The National Institute for Health Research Blood and Transplant Unit (NIHR BTRU) in Donor Health and Genomics at the University of Cambridge.

* Corresponding author

## Contents

# 1 Introduction

In this Supplementary Note we give further information regarding method validation and specifications, and the real data enrichment analysis results.

# 2 Effective number of annotations

For each of the 29 GWAS studies (Online Methods), we downloaded summary statistics, which we used for greedy pruning of the sets of genetic variants per trait (Figure 1). Next, we annotated each variant based on DNaseI hypersensitive site overlap (or LD $r^2 > 0.8$ with such a variant). In order to calculate the effective number of independent annotations we adapted the approach proposed in Galwey, 2009 [17]. Specifically, for each phenotype we took the binary (annotated/not annotated) matrix and calculated the eigenvalues $\lambda_i$ of its correlation matrix. Then the effective number of features was defined as $M_{eff} = (\sum_{i=1}^{M} \sqrt{\lambda_i})^2 / (\sum_{i=1}^{M} \lambda_i)$ , where M = 424 was the total number of annotations (DNaseI hypersensitive sites) used. As a result we found between 182.8 and 194.5 effective numbers of annotations for the 29 traits (Supplementary Table 3) and choose to use the most conservative for a Bonferroni correction of all traits, which gives P-value threshold of $2.6 \times 10^{-4}$ at the 95% significance level. Similarly, for the histone modification and segmentation data we used P-value thresholds of $4.7 \times 10^{-4}$ and $3.3 \times 10^{-5}$ at the 95% significance level, respectively.

# 3 Effect of distance to nearest TSS and number of LD proxies

To investigate the effects of the possible confounding features, we added a correction for each of the features, one at a time, and compared the resulting proportions of significant annotations (DHS data) for each trait to a model that does not account for any of the features. We performed this at the $10^{-8}$ GWAS threshold in all 29 GWAS datasets. We found that in all settings we got fewer enrichments deemed as significant when using a feature correction (Supplementary Figures 1b and 1c). This suggests that a number of enrichments can be explained by our chosen features and therefore not accounting for them can introduce confounding.

# 4 Enrichment of GWAS variants in DNaseI hypersensitive sites

For each of the 29 GWAS studies, we performed enrichment analysis for each 424 ENCODE and Roadmap Epigenomics cell type using GARFIELD such that we calculated odds ratios (OR) at GWAS thresholds ranging from $T < 10^{-1}$ to $T < 10^{-8}$ (in powers of 10) and then tested that enrichment at the four most significant thresholds, from $T < 10^{-5}$ to $T < 10^{-8}$. Supplementary Figure 2 present enrichment wheel plots for each trait. Results highlight sets of traits with cell type specific enrichment as well as ones with an overall enrichment.

# 5 Multiple annotation enrichment estimates for GWAS variants in DNaseI hypersensitive sites

For each of the GWAS studies for which we found multiple enrichments, we performed multiple annotation enrichment analysis for each 424 ENCODE and Roadmap Epigenomics cell type using GARFIELD at GWAS threshold of $T < 10^{-8}$. We first calculated odds ratios (OR) (and 95%CIs) for each annotation on its own, then sorted the annotations in order of significance and finally sequentially added annotations to the model if they significantly improved the model fit (Chi-square p-value $< 0.05$). Supplementary Figure 4 presents the annotations left in the model for each trait at the end of this procedure, and their univariate and multivariate model ORs and 95%CIs.

# 6 Enrichment in 25 state genome segmentations in 127 cell types

Similarly to the DNaseI hypersensitive site data, GARFIELD enrichment analysis was run for each of the 29 phenotypes and each genome segmentation state and cell type. The number of annotations was $25 \times 127 = 3175$ and the significance p-value threshold after Bonferroni correction for the effective number of annotations was $3.3 \times 10^{-5}$. Summary of results for the $T < 10^{-8}$ GWAS significance threshold are shown in Supplementary Figure 5, and state information and cell type information is given in Supplementary Tables 7 and 8, respectively.

# 7 UK10K Project Consortium

Saeed Al Turki [1,2], Carl A. Anderson [1], Richard Anney [3], Dinu Antony [4], Mara Soler Artigas [5], Muhammad Ayub [6], Senduran Bala [1], Jeffrey C. Barrett [1], Ins Barroso [1,7], Phil Beales [4], Jamie Bentham [8], Shoumo Bhattacharya [8], Ewan Birney [9], Douglas Blackwood [10], Martin Bobrow [11], Elena Bochukova [7], Patrick F. Bolton [12,13,14], Rebecca Bounds [7], Chris Boustred [15], Gerome Breen [13,14], Mattia Calissano [16], Keren Carss [1], Ruth Charlton [17], Krishna Chatterjee [7], Lu Chen [1,18], Antonio Ciampi [19], Sebahattin Cirak [16,20], Peter Clapham [1], Gail Clement [21], Guy Coates [1], Massimiliano Cocca [22,23], David A. Collier [14,24], Catherine Cosgrove [8], Tony Cox [1], Nick Craddock [25], Lucy Crooks [1,26], Sarah Curran [12,27,28], David Curtis [29], Allan Daly [1], Petr Danecek [1], Ian N. M. Day [30], Aaron Day-Williams [1,31], Anna Dominiczak [32], Thomas Down [1,33], Yuanping Du [34], Ian Dunham [9], Richard Durbin [1], Sarah Edkins [1], Rosemary Ekong [35], Peter Ellis [1], David M. Evans [36,37], I. Sadaf Farooqi [7], David R. Fitzpatrick [38], Paul Flicek [1,9], James Floyd [1,39], A. Reghan Foley [16], Christopher S. Franklin [1], Marta Futema [40], Louise Gallagher [3], Tom R. Gaunt [36], Matthias Geihs [1], Daniel Geschwind [41], Celia M. T. Greenwood [19,42,43,44], Heather Griffin [45], Detelina Grozeva [11], Xiaosen Guo [34,46], Xueqin Guo [34], Hugh Gurling [47], Deborah Hart [21], Audrey E. Hendricks [1,48], Peter Holmans [25], Bryan Howie [49], Jie Huang [1], Liren Huang [34], Tim Hubbard [1,33], Steve E. Humphries [40], Matthew E. Hurles [1], Pirro Hysi [21], Valentina Iotchkova [1,9], David K. Jackson [1], Yalda Jamshidi [50], Chris Joyce [1], Konrad J. Karczewski [51,52], Jane Kaye [45], Thomas Keane

[1], John P. Kemp [36,37], Karen Kennedy [1,53], Alastair Kent [54], Julia Keogh [7], Farrah Khawaja [55], Margriet van Kogelenberg [1], Anja Kolb-Kokocinski [1], Genevieve Lachance [21], Cordelia Langford [1], Daniel Lawson [56], Irene Lee [57], Monkol Lek [51], Rui Li [42,43,58], Yingrui Li [34], Jieqin Liang [34], Hong Lin [34], Ryan Liu [59], Jouko Lnnqvist[60], Luis R. Lopes [61,62], Margarida Lopes [1,63,64], Daniel G. MacArthur[51,52], Massimo Mangino [21,65], Jonathan Marchini [63,66], Galle Marenne [1], John Maslen [1], Iain Mathieson [67], Shane McCarthy [1], Peter McGuffin [14], Andrew M. McIntosh [10], Andrew G. McKechanie [10,68], Andrew McQuillin [47], Yasin Memari [1], Sarah Metrustry [21], Nicola Migone [69], Josine L. Min [36], Hannah M. Mitchison [4], Alireza Moayyeri [21,70], Andrew Morris [71], James Morris [1], Dawn Muddyman [1], Francesco Muntoni[16], Kate Northstone [36], Michael C. O'Donovan [25], Stephen O'Rahilly [7], Alexandros Onoufriadis [33], Karim Oualkacha [72], Michael J. Owen [25], Aarno Palotie [1,73,74], Kalliope Panoutsopoulou [1], Victoria Parker [7], Jeremy R. Parr [75], Lavinia Paternoster [36], Tiina Paunio [60,76], Felicity Payne [1], Stewart J. Payne [77], John R. B. Perry [21,78], Olli Pietilainen [1,60,73], Vincent Plagnol [79], Rebecca C. Pollitt [80], David J. Porteous [81], Sue Povey [35], Michael A. Quail [1], Lydia Quaye [21], F. Lucy Raymond [11], Karola Rehnstrm [1], J. Brent Richards [19,21,42,43,58], Cheryl K. Ridout [82], Susan Ring [83], Graham R. S. Ritchie [1,9], Nicola Roberts [11], Rachel L. Robinson [17], David B. Savage [7], Peter Scambler [4], Stephan Schiffels [1], Miriam Schmidts [4,84], Nadia Schoenmakers [7], Richard H. Scott [4,85], Robert K. Semple [7], Eva Serra [1], Sally I. Sharp [47], Adam Shaw [86], Hashem A. Shihab [36], So-Youn Shin [1,36], David Skuse [57], Kerrin S. Small [21], Carol Smee [1], Blair H. Smith [87], George Davey Smith [36], Nicole Soranzo [1,18], Lorraine Southam [1,63], Olivera Spasic-Boskovic [11], Timothy D. Spector [21], David St Clair [88], Beate St Pourcain [36,89,90], Jim Stalker [1], Elizabeth Stevens [16], Jianping Sun [19,42], Gabriela Surdulescu [21], Jaana Suvisaari [60], Petros Syrris [61], Ioanna Tachmazidou [1], Rohan Taylor [55], Jing Tian [34], Nicholas J. Timpson [36], Martin D. Tobin [5,91], Ana M. Valdes [21], Anthony M. Vandersteen [92], Parthiban Vijayarangakannan [1], Peter M. Visscher [37,93], Louise V. Wain [5], Klaudia Walter [1], James T. R. Walters [25], Guangbiao Wang [34], Jun Wang [34,46,94,95,96], Yu Wang [34], Kirsten Ward [21], Eleanor Wheeler [1], Tamieka Whyte [16], Hywel J. Williams [25,97], Kathleen A. Williamson [38], Crispian Wilson [11], Scott G. Wilson [21,98,99], Kim Wong [1], ChangJiang Xu [19,42], Jian Yang [37,93], Eleftheria Zeggini [1], Feng Zhang [21], Pingbo Zhang [34], Hou-Feng Zheng [42,43,58]

**Affiliations** 1. The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1HH, Cambridge, UK.
2. Department of Pathology, King Abdulaziz Medical City, P.O. Box 22490, Riyadh 11426, Saudi Arabia.
3. Department of Psychiatry, Trinity Centre for Health Sciences, St James Hospital, James's Street, Dublin 8, Ireland.
4. Genetics and Genomic Medicine and Birth Defects Research Centre, UCL Institute of Child Health, London WC1N 1EH, UK.
5. Departments of Health Sciences and Genetics, University of Leicester, Leicester LE1 7RH, UK.
6. Division of Developmental Disabilities, Department of Psychiatry, Queen's University, Kingston, Ontario N6C 0A7, Canada.
7. University of Cambridge Metabolic Research Laboratories, and NIHR Cambridge Biomed-

ical Research Centre, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK.
8. Department of Cardiovascular Medicine and Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK.
9. European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK.
10. Division of Psychiatry, The University of Edinburgh, Royal Edinburgh Hospital, Edinburgh EH10 5HF, UK.
11. Academic Laboratory of Medical Genetics, Box 238, Lv 6 Addenbrooke's Treatment Centre, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK.
12. Department of Child Psychiatry, Institute of Psychiatry, Psychology and Neuroscience, King's College London, 16 De Crespigny Park, London SE5 8AF, UK.
13. NIHR BRC for Mental Health, Institute of Psychiatry, Psychology and Neuroscience and SLaM NHS Trust, King's College London, 16 De Crespigny Park, London SE5 8AF, UK.
14. MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, Denmark Hill, London SE5 8AF, UK.
15. North East Thames Regional Genetics Service, Great Ormond Street Hospital NHS Foundation Trust, London WC1N 3JH, UK.
16. Dubowitz Neuromuscular Centre, UCL Institute of Child Health & Great Ormond Street Hospital, London WC1N 1EH, UK.
17. Leeds Genetics Laboratory, St James University Hospital, Beckett Street, Leeds LS9 7TF, UK.
18. Department of Haematology, University of Cambridge, Long Road, Cambridge CB2 0PT, UK.
19. Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec H3A 1A2, Canada.
20. Institut fr Humangenetik, Uniklinik Kln, Kerpener Strasse 34, 50931 Kln, Germany.
21. The Department of Twin Research & Genetic Epidemiology, King's College London, St Thomas' Campus, Lambeth Palace Road, London SE1 7EH, UK.
22. Medical Genetics, Institute for Maternal and Child Health IRCCS "Burlo Garofolo", 34100 Trieste, Italy.
23. Department of Medical, Surgical and Health Sciences, University of Trieste, 34100 Trieste, Italy.
24. Lilly Research Laboratories, Eli Lilly & Co. Ltd., Erl Wood Manor, Sunninghill Road, Windlesham GU20 6PH, UK.
25. MRC Centre for Neuropsychiatric Genetics & Genomics, Institute of Psychological Medicine & Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff CF24 4HQ, UK.
26. Sheffield Diagnostic Genetics Service, Sheffield Childrens' NHS Foundation Trust, Western Bank, Sheffield S10 2TH, UK.
27. University of Sussex, Brighton BN1 9RH, UK.
28. Sussex Partnership NHS Foundation Trust, Swandean, Arundel Road, Worthing BN13 3EP, UK.
29. University College London (UCL), UCL Genetics Institute, Darwin Building, Gower Street, London WC1E 6BT, UK.

30. Bristol Genetic Epidemiology Laboratories, School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Clifton, Bristol BS8 2BN, UK.
31. Computational Biology & Genomics, Biogen Idec, 14 Cambridge Center, Cambridge, Massachusetts 02142, USA.
32. Institute of Cardiovascular and Medical Sciences, University of Glasgow, Wolfson Medical School Building, University Avenue, Glasgow, UK G12 8QQ.
33. Department of Medical and Molecular Genetics, Division of Genetics and Molecular Medicine, King's College London School of Medicine, Guy's Hospital, London SE1 9RT, UK.
34. BGI-Shenzhen, Shenzhen 518083, China.
35. University College London (UCL) Department of Genetics, Evolution & Environment (GEE), Gower Street, London WC1E 6BT, UK.
36. MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Clifton, Bristol BS8 2BN, UK.
37. University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland 4102, Australia.
38. MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, at the University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK.
39. The Genome Centre, John Vane Science Centre, Queen Mary, University of London, Charterhouse Square, London EC1M 6BQ, UK.
40. Cardiovascular Genetics, BHF Laboratories, Rayne Building, Institute of Cardiovascular Sciences, University College London, London WC1E 6JJ, UK.
41. UCLA David Geffen School of Medicine, Los Angeles, California 90095, USA.
42. Lady Davis Institute, Jewish General Hospital, Montreal, Quebec H3T 1E2, Canada.
43. Department of Human Genetics, McGill University, Montreal, Quebec H3A 1B1, Canada.
44. Department of Oncology, McGill University, Montreal, Quebec H2W 1S6, Canada.
45. HeLEX - Centre for Health, Law and Emerging Technologies, Nuffield Department of Population Health, University of Oxford, Old Road Campus, Oxford OX3 7LF, UK.
46. Department of Biology, University of Copenhagen, Ole Maales Vej 5, DK-2200 Copenhagen, Denmark.
47. University College London (UCL), Molecular Psychiatry Laboratory, Division of Psychiatry, Gower Street, London WC1E 6BT, UK.
48. Department of Mathematical and Statistical Sciences, University of Colorado, Denver, Colorado 80204, USA.
49. Adaptive Biotechnologies Corporation, Seattle, Washington 98102, USA.
50. Human Genetics Research Centre, St George's University of London, London SW17 0RE, UK.
51. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.
52. Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA.
53. National Cancer Research Institute, Angel Building, 407 St John Street, London EC1V 4AD, UK.
54. Genetic Alliance UK, 4D Leroy House, 436 Essex Road, London N1 3QP, UK.
55. SW Thames Regional Genetics Lab, St George's University, Cranmer Terrace, London

SW17 0RE, UK.
56. Schools of Mathematics and Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Clifton, Bristol BS8 2BN, UK.
57. Behavioural and Brain Sciences Unit, UCL Institute of Child Health, London WC1N 1EH, UK.
58. Department of Medicine, Jewish General Hospital, McGill University, Montreal, Quebec H3A 1B1, Canada.
59. BGI-Europe, London EC2M 4YE, UK.
60. National Institute for Health and Welfare (THL), Helsinki FI-00271, Finland.
61. Institute of Cardiovascular Science, University College London, Gower Street, London WC1E 6BT, UK.
62. Cardiovascular Centre of the University of Lisbon, Faculty of Medicine, University of Lisbon, Avenida Professor Egas Moniz,1649-028 Lisbon, Portugal.
63. Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK.
64. Illumina Cambridge Ltd, Chesterford Research Park, Cambridge CB10 1XL, UK.
65. National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London SE1 9RT, UK.
66. Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK.
67. Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.
68. The Patrick Wild Centre, The University of Edinburgh, Edinburgh EH10 5HF, UK.
69. Department of Medical Sciences, University of Torino, 10124 Torino, Italy.
70. Institute of Health Informatics, Farr Institute of Health Informatics Research, University College London (UCL), 222 Euston Road, London NW1 2DA, UK.
71. Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, 9 Little France Road, Edinburgh, UK EH16 4UX.
72. Department of Mathematics, Universit de Qubec  Montral, Montral, Qubec H3C 3P8, Canada.
73. Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki FI-00014, Finland.
74. Program in Medical and Population Genetics and Genetic Analysis Platform, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02132, USA.
75. Institute of Neuroscience, Henry Wellcome Building for Neuroecology, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, UK.
76. University of Helsinki, Department of Psychiatry, Helsinki FI-00014, Finland.
77. North West Thames Regional Genetics Service, Kennedy-Galton Centre, Northwick Park Hospital, Watford Road, Harrow HA1 3UJ, UK.
78. MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Box 285, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK.
79. University College London (UCL) Genetics Institute (UGI) Gower Street, London WC1E 6BT, UK.
80. Connective Tissue Disorders Service, Sheffield Diagnostic Genetics Service, Sheffield Children's NHS Foundation Trust, Western Bank, Sheffield S10 2TH, UK.
81. Centre for Genomic and Experimental Medicine, Institute of Genetics and Experimental Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, UK

EH4 2XU.
82. Molecular Genetics, Viapath at Guy's Hospital, London SE1 9RT, UK.
83. ALSPAC & School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Clifton, Bristol BS8 2BN, UK.
84. Human Genetics Department, Radboudumc and Radboud Institute for Molecular Life Sciences (RIMLS), Geert Grooteplein 25, Nijmegen 6525 HP, The Netherlands.
85. Department of Clinical Genetics, Great Ormond Street Hospital, London, WC1N 3JH, UK.
86. Clinical Genetics, Guy's & St Thomas' NHS Foundation Trust, London SE1 9RT, UK.
87. Mackenzie Building, Kirsty Semple Way, Ninewells Hospital and Medical School, Dundee, UK DD2 4RB.
88. Institute of Medical Sciences, University of Aberdeen, Aberdeen AB25 2ZD, UK.
89. School of Oral and Dental Sciences, University of Bristol, Lower Maudlin Street, Bristol BS1 2LY, UK.
90. School of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol BS8 1TU, UK.
91. National Institute for Health Research (NIHR) Leicester Respiratory Biomedical Research Unit, Glenfield Hospital, Leicester LE3 9QP, UK.
92. Maritime Medical Genetics Service, 5850/5980 University Avenue, PO Box 9700, Halifax, Nova Scotia B3K 6R8, Canada.
93. Queensland Brain Institute, University of Queensland, Brisbane, Queensland 4072, Australia.
94. Princess Al Jawhara Albrahim Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, P.O. Box 80200, Jeddah 21589, Saudi Arabia.
95. Macau University of Science and Technology, Avenida Wai long, Taipa, Macau 999078, China.
96. Department of Medicine and State Key Laboratory of Pharmaceutical Biotechnology, University of Hong Kong, 21 Sassoon Road, Hong Kong.
97. The Centre for Translational Omics - GOSgene, UCL Institute of Child Health, London WC1N 1EH, UK.
98. School of Medicine and Pharmacology, University of Western Australia, Perth, Western Australia 6009, Australia.
99. Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Nedlands, Western Australia 6009, Australia.