EMBO Molecular Medicine

The human variome: genomic and epigenomic diversity

Work from the HUGO Pan-Asian Single Nucleotide Polymorphism (SNP) consortium has mapped human genetic diversity in different subpopulations in Asia (Abdulla et al, 2009). Now this highlights the question whether disease-SNP associations identified by genome-wide association studies (GWAS) in one subpopulation can be applied to other subpopulations or whether separate GWAS should be considered for different subpopulations because of their different SNP distributions (Haiman & Stram, 2010).

DNA methylation is an epigenetic mark in the mammalian genome. 'Epigenetic diversity' encompasses variation in DNA methylation and like 'genetic diversity', 'epigenetic diversity' may also covary with phenotype (Feinberg et al, 2010). Mechanistically, disease-associated SNPs may be found in DNA regulatory regions where the genetic variation affects cognate binding of transcription factor complexes, effecting on the expression of disease-relevant genes (Harismendy et al, 2011). Similarly, variation in the methylation of a DNA regulatory region (either increased or decreased methylation) can also alter the binding of transcription factor complexes and regulate both distal and proximal gene expression (Choy et al, 2010; Jones & Takai, 2001; Phillips & Corces, 2009; Weaver et al, 2004). The availability of genome-wide maps of DNA methylation (Choy et al, 2010; Down et al, 2008; Lister et al, 2009) led us to examine whether a relationship exists between DNA methylation and regions of heterogeneous genetic variation.

The Fixation Index (Fst) is a measure of fixation rate of SNP between two subpopulations. SNPs with high Fst are ones with more heterogeneity in a subpopulation compared to another. Using DNA methylation maps that were generated from individuals of European ancestry, we made the interesting discovery that high Fst-SNPs (Duan et al, 2008) ($N = 8751$) from European ancestry showed an aggregate methylation density pattern of modulation centred on SNPs compared to flanks of the SNPs (Fig 1). Methylation maps from Europeans showed density modulation only at sites of European high Fst-SNPs (i.e. high Fst-SNPs for European *vs*. African + European *vs*. Asian), but not sites of Asian or African high Fst-SNPs (i.e. high Fst-SNPs for Asian *vs*. European; or African *vs*. European, Fig 1). Similarly aggregate methylation density modulation was also absent in a negative control set of randomly selected SNPs ($N = 8000$). On the basis that increased DNA methylation may be a

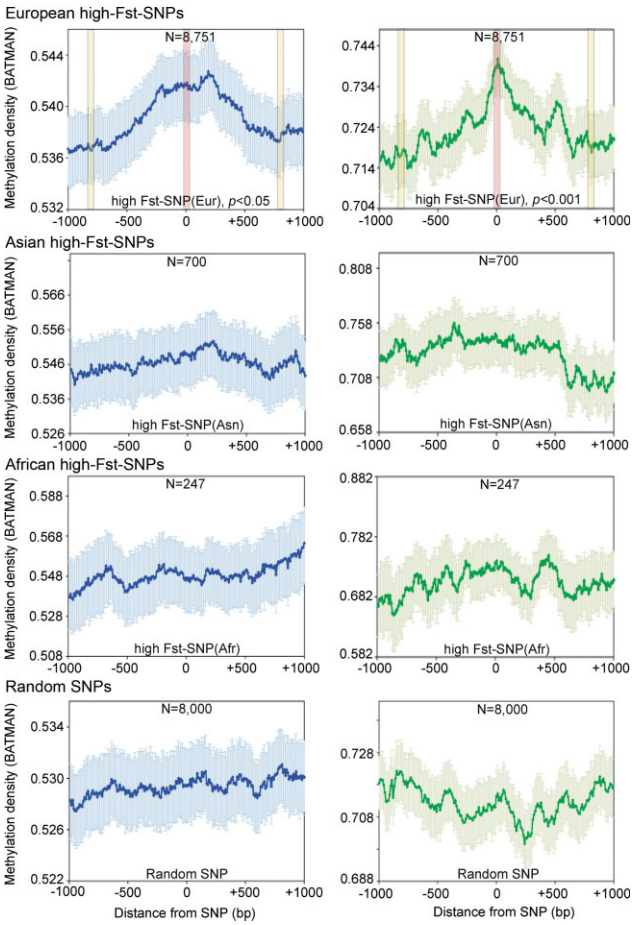

Figure 1. Methylation density across sites of SNPs. Methylation maps from hearts (blue) and sperm (green) from Europeans were used and aggregate methylation density was plotted against sites −1000 to +1000 bp of European, Asian, African high-Fst-SNPs, and random SNPs. European methylation density showed a statistically significant pattern of modulation only in European high Fst-SNPs but not high Fst-SNPs of other subpopulations (one-way ANOVA for methylation density at red and yellow bars, $p < 0.05$ for hearts and $p < 0.001$ for sperm). Aggregate methylation density (BATMAN score) was generated and plotted using algorithms previously described (Choy et al, 2010; Down et al, 2008). MeDIP-seq (Methylated DNA-immunoprecipitation followed by high-throughput sequencing) data were generated from four human hearts and spermatozoa and analyzed using a Bayesian deconvolution strategy (BATMAN score) (Choy et al, 2010; Down et al, 2008). BATMAN scores from four healthy human hearts were averaged using Perl script written by HGG and MKC as described in Choy et al, (2010). Average plots of methylation densities were calculated using an algorithm found in a public resource (Kolasinska-Zwierz et al, 2009).

means of regulating transcription factor complex binding and therefore mark genomic regions that are important to control by methylation (Choy et al, 2010; Jones & Takai, 2001), our findings support the notion that sites of heterogeneous genetic variation in one subpopulation are functionally relevant to the corresponding subpopulation, but sites of heterogeneous genetic variation of a different subpopulation, are not.

This is consistent with the need to consider subpopulation differences in genetic variation when studying disease-SNP association. By extension, we propose that the epigenomes of subpopulations should also be considered as the International Human Epigenome Consortium gears up to generate 1000 reference epigenomes (Beck, 2010). Subpopulation differences will be important when drawing conclusions from associations between disease and the human variome.

The authors declare that they have no conflict of interest.

References

Abdulla MA, Ahmed I, Assawamakin A, Bhak J, Brahmachari SK, Calacal GC, Chaurasia A, Chen CH, Chen J, Chen YT, *et al* (2009) Mapping human genetic diversity in Asia. Science 326: 1541-1545

Beck S (2010) Taking the measure of the methylome. Nat Biotechnol 28: 1026-1028

Choy MK, Movassagh M, Goh HG, Bennett MR, Down TA, Foo RS (2010) Genome-wide conserved consensus transcription factor binding motifs are hyper-methylated. BMC Genomics 11: 519

Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Kulesha E, Graf S, Johnson N, Herrero J, Tomazou EM, *et al* (2008) A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. Nat Biotechnol 26: 779-785

Duan S, Zhang W, Cox NJ, Dolan ME (2008) FstSNP-HapMap3: a database of SNPs with high population differentiation for HapMap3. Bioinformation 3: 139-141

Feinberg AP, Irizarry RA, Fradin D, Aryee MJ, Murakami P, Aspelund T, Eiriksdottir G, Harris TB, Launer L, Gudnason V, *et al* (2010) Personalized epigenomic signatures that are stable over time and covary with body mass index. Sci Transl Med 2: 49ra67

Haiman CA, Stram DO (2010) Exploring genetic susceptibility to cancer in diverse populations. Curr Opin Genet Dev 20: 330-335

Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, Ren B, Fu XD, Topol EJ, Rosenfeld MG, *et al* (2011) 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. Nature 470: 264-268

Jones PA, Takai D (2001) The role of DNA methylation in mammalian epigenetics. Science 293: 1068-1070

Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. Nat Genet 41: 376-381

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, *et al* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462: 315-322

Phillips JE, Corces VG (2009) CTCF: master weaver of the genome. Cell 137: 1194-1211

Weaver IC, Cervoni N, Champagne FA, D'Alessio AC, Sharma S, Seckl JR, Dymov S, Szyf M, Meaney MJ (2004) Epigenetic programming by maternal behavior. Nat Neurosci 7: 847-854

Munkit Choy, Mehregan Movassagh, Roger Foo in Department of Medicine, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK
Tel: +44 122 3331504, Fax: +44 122 3331505
E-mail: rsyf2@cam.ac.uk

DOI 10.1002/emmm.201100164

