Vol.2011-BIO-25 No.39
2011/6/24

Tissue specific methylation and genotype

Ryoichi Kinoshita[†1] and Y-h. Taguchi[†1]

It is unclear how genotype affects methylation. In this paper, we have applied principal component analysis to both genotype and methylation on tumor, adjacent normal tissue, and blood DNA from 30 patients. Dominant pattern turns out to be upregulation from blood to tumor through tissued for both genotype and methylation. Overlap between significantly upregulating genotype and methylized genes are significantly overlaped. Thus, at least, for some regions, genotype correlates to methylation.

1. Introduction

It is unclear how genotype affects methylation. Recently Yang et al[1)] investigated that relationship between genotype and methylation and concluded that genotype affects methylation although genotype itself lacks sufficient tissue specificity using principal component analysis (PCA). In this paper, we have applied PCA to genotype and methylation patterns on tumor, adjacent normal tissue, and blood DNA from 30 patients in the different way. Then we have found that genotypes frequency of which upregulates from blood to tumor through normal tissue are significantly overlapped with those methylation of which levels from blood to tumor through normal tissue is upregulated. Thus, in contrast to Yang et al's results, genotype is directly related to methylation.

2. Materials and Methods

2.1 Genotype and methylation pattern

Both genotype and methylation pattern are obtained from GEO by the accession number of GSE20123[1)]. In this paper, we have employed results from Nsp chip only for simplicity.

†1 Department of Physics, Chuo University, Tokyo, Japan

2.2 PCA analysis

PCA is applied to genotype frequency and DNA methylation pattern of individual probes. Then all probes are embedded into two dimensional space spanning from the 2nd and 3rd principal component for gene expression and DNA methylaion separately.

2.3 P-values

P-values between three kinds of tissues are computed for each probes by t-test. For this, we have used t.test module in R[2)].

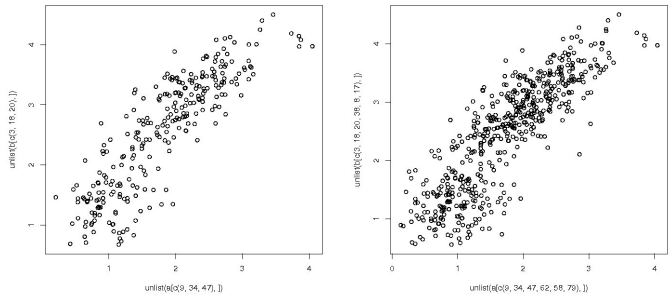
P-values for overlap between top upregulating probes between genotype frequency and methylation is computed by binary distribution by using pbinom module in R[2)].

3. Results

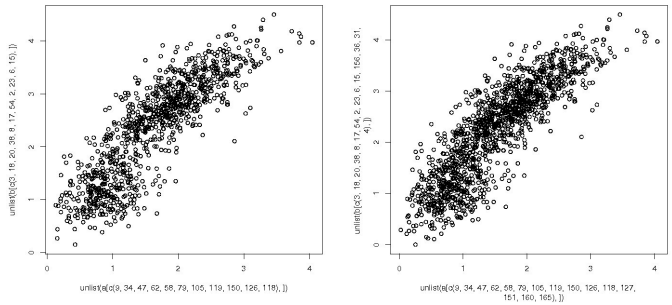
After applying PCA to both genotype frequency and methylation pattern, we have found 2nd PCA component correspond to upregulation from blood(B) to tumor(T) through normal(N) tissue. The first principal component lacks tissue specificity at all, thus is ignored. Then we have picked up top 300 probes for genotype frequency and methylation due to 2nd PCA component. Since each probe has genotype frequency and methylation patterns of 90 samples which consist of 30 B samples, 30 N samples, and 30 T samples, we have applied t-test of three kinds of pairs of sample sets, i.e., B vs T, B vs N, and T vs N. 298(256) out of 300 probes has at least one P-values satisfying $P < 0.05/N$ for gene expression(methyaltion) patterns, where $N = 262339$ is number of probes in array. Thus, PCA successfully picked up probes whose genotype frequency and methylation patterns are significantly different between tissues.

Next, we have compared these sets with each other. It turns out that top 300 genotype frequency set has siginificant number of common probes with those of methylation ($P < 1 \times 10^{-22}$, see Table 1). This means, genotype is directly related to methylation. It is opposite to what Yang et al[1)] has insisted.

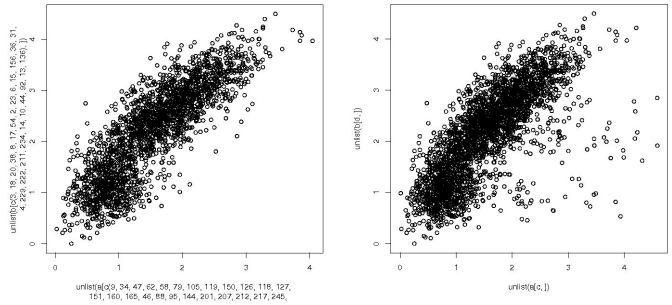
We have also compared genotype freqnency with methylation directly. We have picked up probes for which significance of difference among B, N and T is same and computed correlation coefficients of these probes (see Fig. 1). It is clear that correlation is very strict.



$n = 50, n'_{common} = 3, r = 0.87,\ n = 100, n'_{common} = 6, r = 0.85$



$n = 150, n'_{common} = 11, r = 0.85,\ n = 200, n'_{common} = 15, r = 0.86,$



$n = 250, n'_{common} = 25, r = 0.84,\ n = 300, n'_{common} = 33, r = 0.75.$

Fig. 1 Comparison between methylation(horizontal) and genotype(vertical) of each probes for top n probes. n'_{common} is the number of common probes, r is correlation coefficient.

Table 1 n_{common} is the number of common probes in both of top n probes for genotype/methylation pattern. P-values are always less than 10^{-22}.

n	50	100	150	200	250	300
n_{common}	6	11	22	30	46	68

Table 2 Probes selected. Probes listed at row n is ranked between $n - 50$ and n.

Top n	probes
50	SNP_A-1825620, SNP_A-2309865, SNP_A-4233167
100	SNP_A-1984943, SNP_A-2172952, SNP_A-2234716
150	SNP_A-1988914, SNP_A-2040111, SNP_A-2089983, SNP_A-4195285, SNP_A-4199352
200	SNP_A-2276203, SNP_A-4196078, SNP_A-4226834, SNP_A-4229534
250	SNP_A-1980533, SNP_A-1886593, SNP_A-2042678, SNP_A-2143521, SNP_A-1880907, SNP_A-1989613, SNP_A-2142865, SNP_A-4193660, SNP_A-1845324, SNP_A-1852621
300	SNP_A-2053247, SNP_A-1911642, SNP_A-2221049, SNP_A-4213049, SNP_A-4228665, SNP_A-4236336, SNP_A-2287632, SNP_A-2043441

In Table 2, we have listed probes selected. Although we have checked them at affimetrix site[⋆1], we could not find anything biological meaningful. This must be clarified in the future.

4. Conclusion

In this paper, we have shown gene body methylation is positively correlated to genotype. Further investigation will be needed what causes this effect.

5. Acknowledgement

This work was supported by KAKENHI 23300357.

References

1) Yang, H.H., Hu, N., Wang, C., Ding, T., Dunn, B.K., Goldstein, A.M., Taylor, P.R., Lee, M.P.: Influence of genetic background and tissue types on global DNA methylation patterns, *PLoS One* vol.5, e9355, (2010).
2) R Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org (2009)

⋆1 Check "Genotyping" at http://www.affymetrix.com/analysis/netaffx with selecting "Genome-wide SNP 5.0"