

Table 2. Association Analysis of Intracranial Aneurysm With 9p21 Single Nucleotide Polymorphisms

SNP	Position*	Allele†	Cases				Control				Hardy-Weinberg Equilibrium P	OR (95% CI)	Trend Test P	Nonadditive P ‡
			1/1	1/2	2/2	Frequency‡	1/1	1/2	2/2	Freq.‡				
rs1333040	22073404	T	517	379	85	72.0	287	322	90	64.1	1.00	1.43 (1.24–1.66)	1.5×10^{-6}	0.36
rs2891168	22088619	G	267	451	226	52.2	140	302	207	44.8	0.13	1.32 (1.15–1.52)	7.4×10^{-5}	0.74
rs2383207	22105959	G	459	390	89	69.7	272	288	101	62.9	0.10	1.34 (1.16–1.55)	8.3×10^{-5}	0.35
rs10757278	22114477	G	300	457	220	54.1	160	329	206	46.7	0.20	1.33 (1.16–1.52)	4.1×10^{-5}	0.85

*The position in National Center for Biotechnology Information Genome Build 36.

†The putative susceptibility allele is listed. Genotype counts are expressed relative to the putative susceptibility allele, where allele 1 is the putative susceptibility allele and allele 2 is the nonsusceptible allele.

‡Frequency of the putative susceptibility allele.

§The exact test of departure from Hardy-Weinberg equilibrium in control samples.

¶Nonadditive effects were modeled in the context of logistic regression analysis.

SNP indicates single nucleotide polymorphism.

gender and age (<50, 50–59, 60–69, ≥ 70 years) to address possible confounding. We investigated whether the magnitude of the association between the 9p21 SNP and IA was modified by the risk factors. The joint association of the 9p21 genotype and each risk factor with IA was evaluated. Six categories were formed by combining the 9p21 genotype (AA, Aa, aa) and either smoking habit (ever, never) or history of hypertension (yes, no). Furthermore, interactions between 9p21 and risk factor were investigated by adding a product term into the multivariate logistic regression model to assess whether deviations from multiplicative joint effects were statistically significant (or to examine whether the joint effects were significantly greater or smaller than expected).

To evaluate whether the association between the 9p21 SNP and risk of harboring IA varied according to subphenotypes, we examined the association between the 9p21 SNP and IA according to subgroups stratified by history of SAH (ruptured or unruptured), multiplicity of IA (multiple or single), and location of the IA. When analyzing the history of SAH and location of IA, we excluded patients with multiple IAs because some patients with multiple IAs simultaneously had different types of aneurysms according to the SAH status or aneurysm location (eg, patients with ruptured and unruptured IAs). Thus, 713 patients harboring a single IA were analyzed.

For the subgroup analysis regarding location of IA, the definition of 2 sets of case-subgroups was as follows: (1) patients harboring IA at 3 common sites (anterior communicating artery [AcomA], middle cerebral artery [MCA], and posterior communicating artery [PcomA]), which we defined as the site where the number of patients was >100 ; and (2) patients grouped according to the International Study of Unruptured Intracranial Aneurysms⁵: (i) AC/MC/IC, which signify aneurysms in the AcomA, the anterior cerebral artery (ACA), the MCA, or the internal carotid artery (ICA) (not the cavernous portion) and (ii) Post-Pcom, meaning the vertebrobasilar, the posterior cerebral arterial system, or the PcomA. Distribution of the IA subphenotypes is seen in Table 1.

To estimate subphenotype-specific ORs, the frequency of the 9p21 variant in each case-subgroup (eg, patients with ruptured or unruptured IA) was compared with the frequency in controls using a polytomous logistic regression. To assess whether the estimated subphenotype-specific ORs were significantly different, the frequencies of the 9p21 variant were compared between case-subgroups by dichotomous logistic regression (the case-subgroup heterogeneity test).¹⁶ For these subgroup analyses, the effects of gender and age were adjusted. When assessing the subgroup-specific effects stratified by SAH status, the case subgroup heterogeneity test with adjustment for IA site (6 categories: AcomA, ACA, MCA, ICA, PcomA, the posterior cerebral arterial system) was also implemented to keep out possible confounding effects attributable to complex relationships among IA subphenotypes. Similarly, when assessing the subgroup-specific effects stratified by IA sites, the effects of SAH status were adjusted (2 categories: ruptured, unruptured). All the logistic regression analyses were performed by SAS (SAS Institute).

Results

The results of genotyping and association analysis for 4 SNPs on 9p21 are shown in Table 2. For the SNPs, the missing genotype rates were small (at most, 5.2% for rs2891168) and the genotypes for all SNPs were in the Hardy-Weinberg equilibrium in controls ($P>0.05$). All the SNPs showed highly significant associations with IA by means of the trend test ($P=1.5 \times 10^{-6}$ for rs1333040; $P=7.4 \times 10^{-5}$ for rs2891168; $P=8.3 \times 10^{-5}$ for rs2383207; $P=4.1 \times 10^{-5}$ for rs10757278). The effect of non-additivity was not significant for any SNP ($P>0.05$); therefore, the multiplicative model was selected as an appropriate genetic model and applied in subsequent logistic regression analyses.

For the stepwise logistic regression analysis, the missing genotype imputation was implemented. The estimated R^2 values were >0.98 for all the SNPs, indicating that accuracy of the imputation was considerably high. The stepwise logistic regression analysis identified rs1333040 as an independent predictor of IA among the 4 SNPs. The observed associations of the other 3 SNPs could be a result of linkage disequilibrium with rs1333040. Haplotype analysis reinforces this inference. All but 1 of 4 common haplotypes bearing T allele of rs1333040 (ie, H3, H5, and H6) was significantly associated with IA, whereas haplotype H2 bearing the C allele of rs1333040 and the risk alleles of the other 3 SNPs did not show significant association (Table 3). Thus, only rs1333040 was considered in further analyses.

The ORs obtained from both univariate and multivariate logistic regression analyses are shown in Table 4. The univariate OR for rs1333040 was 1.43 (95% CI, 1.24–1.66), which was similar to the OR of 1.44 (95% CI, 1.22–1.72) after the adjustment in the multivariate model that included gender, age, smoking, and hypertension. This result indicates that the association of this SNP is independent of the risk factors. However, the ORs for smoking habit and history of hypertension in the univariate model (1.33 [95% CI, 1.08–1.64] for smoking and 1.67 [95% CI, 1.36–2.06] for hypertension) were largely different from those in the multivariate model (2.32 [95% CI, 1.75–3.08] for smoking and 2.25 [95% CI, 1.78–2.86] for hypertension). This may be attributable to the correlations among the factors; smoking correlated negatively with age and female gender, and hypertension correlated positively with age (trend test $P<0.0001$).

Table 3. Haplotype Analysis of 9p21 Single Nucleotide Polymorphisms*

Haplotype†	SNP				Frequency (%)	OR (95% CI)‡	P
	rs1333040	rs2891168	rs2383207	rs10757278			
H1	C	A	A	A	25.7	Reference	...
H2	C	G	G	G	4.2	1.13 (0.79–1.61)	0.52
H3	T	A	A	A	6.8	1.43 (1.07–1.93)	0.018
H4	T	A	G	A	14.6	1.23 (0.98–1.53)	0.073
H5	T	A	G	G	2.4	1.62 (1.01–2.59)	0.044
H6	T	G	G	G	43.8	1.56 (1.32–1.85)	2.9×10 ⁻⁷

*The haplotype phasing was performed using Beagle software.¹⁵

†Six common haplotypes whose frequencies exceeded 1% were analyzed.

‡OR and 95% CI for each haplotype was estimated by setting haplotype H1, bearing no putative susceptibility allele at every SNP site, as the reference category.

The omnibus haplotype test with the degree of freedom of 5 was also performed. The likelihood ratio χ^2 statistic was 29.017 and the resulting *P* value was 2.2×10⁻⁵.

The result of the joint association of rs1333040 and each risk factor is displayed in Supplemental Figure II available online at <http://stroke.ahajournals.org>, which shows that the effect of environmental exposure (smoking or hypertension) did not differ among subjects stratified by genotypes on a multiplicative (log-odds) scale and vice versa. Furthermore, the tests for deviations from multiplicative joint effects were not significant; the ORs for the genotype–smoking and genotype–hypertension interactions were 1.08 (95% CI, 0.76–1.53; *P*=0.67) and 1.22 (95% CI, 0.86–1.72; *P*=0.27), respectively.

The associations between rs1333040 and IA subdivided by aneurysmal subphenotypes were examined in the context of the polytomous logistic regression analyses. The subphenotype-specific ORs are shown in the Figure. The rs1333040 association was stronger for multiple IAs (OR, 1.57; 95% CI, 1.21–2.03) than for single IA (OR, 1.36; 95% CI, 1.15–1.61). However, the case-subgroup heterogeneity test did not show significant difference in OR between multiple and single IAs (*P*=0.29).

The subgroup analysis stratified by SAH status showed that rs1333040 was associated with ruptured IA (OR, 1.42; 95% CI, 1.18–1.71) and unruptured IA (OR, 1.27; 95% CI, 1.00–1.62). The difference in OR between ruptured and unruptured IAs was not significant in terms of the case-subgroup heterogeneity test (*P*=0.42). This was confirmed by the case-subgroup heterogeneity test with adjustment for IA site (*P*=0.69).

When stratifying patients with a single IA by 3 common sites, the association was strongest for PcomA (OR, 1.69; 95% CI, 1.26–2.26), secondly for MCA (OR, 1.36; 95% CI,

1.07–1.72), and not significant for AcomA (OR, 1.22; 95% CI, 0.96–1.57). Then, the difference in OR between PcomA and AcomA was suggestive by the case-subgroup heterogeneity test (*P*=0.067). However, the case-subgroup heterogeneity tests were not significant (*P*=0.50 for AcomA vs MCA and *P*=0.23 for MCA vs PcomA). When dichotomizing IA sites according to the grouping in the International Study of Unruptured Intracranial Aneurysm, the association was stronger for Post-Pcomm (OR, 1.73; 95% CI, 1.32–2.26) than for AC/MC/IC (OR, 1.28; 95% CI, 1.07–1.53). We found the difference between these ORs was statistically significant in the case-subgroup heterogeneity test (*P*=0.032). To consider a possible relationship between location of IA and SAH status, the difference among the site-specific ORs was also assessed with adjustment for SAH status. The resulting *P* values for the case-subgroup heterogeneity tests were *P*=0.055 for AcomA vs PcomA and *P*=0.031 for Post-Pcomm vs AC/MC/IC. These results indicate that the 9p21 SNP is more strongly associated with IA in the posterior circulation or PcomA than that in the anterior circulation.

Discussion

We selected 4 9p21 SNPs to capture SNPs showing strong association with several arterial diseases in previous studies and then confirmed that the SNPs were associated with IA in Japanese. Among the 4 SNPs, rs1333040 showed the strongest evidence of association. The stepwise logistic regression analysis and the haplotype analysis reveal that the sequence variant tagged by the T allele of rs1333040 can explain the observed association with IA. Recently, the association between the 9p21 locus and IA was confirmed in a white population that was independent from populations analyzed in the original genome-wide association study.¹⁷

We found that the association was independent of established risk factors (gender, age, smoking, and hypertension). The presence of missing data on smoking and hypertension was a limitation of our study. However, similar findings were observed in case-control studies of IA and coronary artery disease; the effect of the 9p21 locus on risk of IA and coronary artery disease was not altered by adjustment for the risk factors.^{12,17,18} This may support the validity of our result.

Table 4. Odds Ratios for rs1333040 and Clinical Risk Factors With Intracranial Aneurysm

Characteristics, Contrast	OR (95% CI)	
	Univariate	Multivariate*
rs1333040, T vs C	1.43 (1.24–1.66)	1.44 (1.22–1.72)
Smoking habit, ever vs never	1.33 (1.08–1.64)	2.32 (1.75–3.08)
History of hypertension, yes vs no	1.67 (1.36–2.06)	2.25 (1.78–2.86)

*Multivariate logistic regression model that included all these predictors with adjustment for gender and age (<50, 50–59, 60–69, ≥70). In the multivariate logistic regression analysis, we used data from 768 cases and 626 controls with complete information for all the variables.

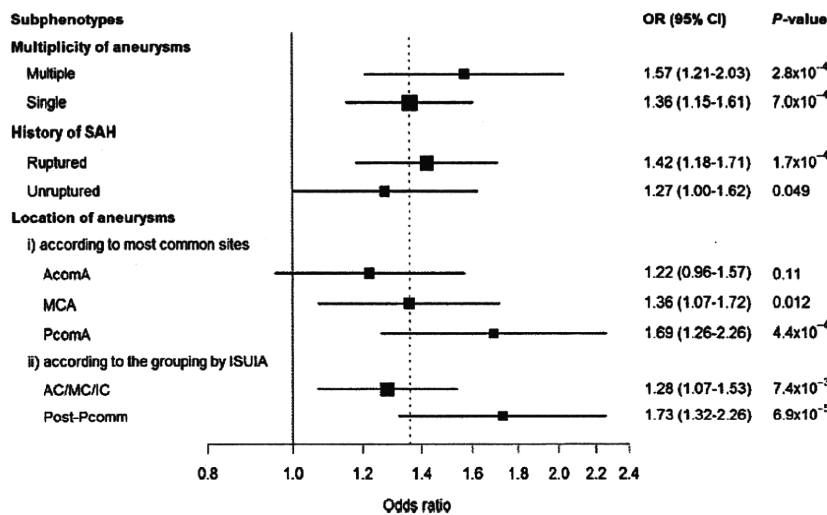


Figure. Forest plots displaying ORs and 95% CIs for the association between rs1333040 and intracranial aneurysm (IA) stratified by subphenotypes. See definitions of abbreviations for the location of IA and the grouping of IA sites in Table 1. The vertical dotted line shows OR for single IA patients because we restricted analyses to patients with single IA for subgroup analyses regarding history of subarachnoid hemorrhage and location of IA. All the ORs were estimated with adjustment for gender and age.

The most important findings of this study are that rs1333040 was more strongly associated with IA of the PcomA and the posterior circulation than with IA in the anterior circulation. In terms of common sites, we found that the T allele of rs1333040 was associated with a 1.69-fold increased risk of IA in PcomA relative to the low-risk allele C, whereas the increased risk of IA in AcomA was only 1.22-fold.

To our knowledge, this is the first study showing that a common genetic variant with widely replicated evidence is associated with site distribution of IA. Epidemiological studies reinforce the hypothesis that genetic components influence site distribution of IA. There were considerably high concordance rates of IAs at identical or mirror-image sites in twins (70%)¹⁹ and in sibling pairs (69%),²⁰ whereas the concordance rate in randomly selected pairs of patients was only 21%.¹⁹ These findings suggest that the 9p21 variant is one of the genetic factors predisposing to IAs in specific locations. Recently, Lindner et al²¹ demonstrated that the effects of established risk factors (gender, age, and alcohol intake) were different according to the site of aneurysms. This finding suggests that there is a possibility that the effects of genetic risk factors may also differ according to the site of aneurysms. The natural history of IAs in the anterior circulation and the posterior circulation and the PcomA are different. The posterior group of aneurysms is associated with a higher risk of rupture and poorer surgical and endovascular outcomes compared to the anterior circulation aneurysms.^{5,6} In the future, functional analyses focusing on differential gene expressions of aneurysm dome tissue subdivided by arterial sites may identify susceptibility genes and polymorphisms responsible for the formation of each subphenotype of IA and may help further understand the pathogenesis of this complex disease.

The effect of 9p21 SNP was not different when stratifying IA patients by SAH status. This finding suggests that the 9p21 risk locus may be involved in the formation of IA rather than in the precipitation of the aneurysm rupture.

It is known that subphenotypes of IAs show common characteristics, eg, the risk of aneurysm rupture is associated with aneurysm size and location.^{5,6} It is important to rule out possible confounding effects within the groups of IA subphe-

notypes. The analytical methods used in this study worked well for subgroup analyses regarding history of SAH and location of IA. The analysis for multiplicity of IA is more complex because of the presence of several subphenotypes in 1 patient eg, aneurysm at the AcomA and the MCA. Further sophisticated and large-scale analyses are needed to identify susceptibility genes of multiplicity of IA.

Some limitations of our study should be noted. All the current analyses are based on statistical analyses; therefore, functional analysis of rs1333040 or other SNP that is in linkage disequilibrium with rs1333040 is necessary. Although the diagnostic performance of the noninvasive imaging methods for the evaluation of IAs has been significantly improved, the use of magnetic resonance angiography and computed tomography angiography may be subject to misclassification of IA subphenotypes as compared to digital subtraction angiography and surgical findings. Subgroup analyses are subject to inflated false-positive rates attributable to multiple testing.^{22,23} The main conclusion of this study is based on 6 case-subgroup heterogeneity tests to assess whether the estimated subphenotype-specific ORs were significantly different. For simplicity, when we assume the tests are unrelated, 6 tests may produce a 26% chance of observing at least 1 significant result at the significance level of 0.05. Therefore, independent replication studies and meta-analyses integrating results from original and subsequent replication studies are indispensable to establish the credibility of the findings.^{22,23} We dichotomized aneurysm sites according to the grouping in the International Study of Unruptured Intracranial Aneurysm,⁵ but there may be controversy regarding the grouping of the PcomA aneurysms because the PcomA derives from the ICA.²⁴ Therefore, both of the findings when IA patients are stratified according to the International Study of Unruptured Intracranial Aneurysm grouping and according to the common site will be corroborated.

Conclusion

One of the difficulties in genetic dissection of IA may be attributable to possible phenotypic heterogeneity.^{1,7} In most association studies, patients with different aneurysmal pheno-

types were grouped together.⁷ If different predisposition underlies different subphenotypes of IA, then reducing phenotypic heterogeneity by restricting to specific subphenotype would be effective. We assessed whether the magnitude of the 9p21 association varied by aneurysmal subphenotypes. Our results of the 9p21 variation being associated with location of IA suggest that a genetic dissection of subphenotypes of IA is important and feasible. One of the most important aspects of IAs is whether they will rupture. The analytical methods implemented in this study may be useful to identify a genetic factor that allows prediction of aneurysmal rupture.

Acknowledgments

The authors thank all the study participants and supporting medical staff for making this study possible. The authors are grateful to Drs. Hideaki Onda, Taku Yoneyama, Hiroyuki Akagawa, Makiko Funamizu, and Midori Yamamoto for their contribution to sample and data collection. The authors thank Hiromi Kamura for her technical support.

Sources of Funding

This work was supported by KAKENHI (Grant-in-Aid for Scientific Research) on Priority Areas "Applied Genomics" from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

Disclosure

None.

References

- Nahed BV, Bydon M, Ozturk AK, Bilguvar K, Bayrakli F, Gunel M. Genetics of intracranial aneurysms. *Neurosurgery*. 2007;60:213–225.
- Bilguvar K, Yasuno K, Niemela M, Ruigrok YM, von Und Zu Fraunberg M, van Duijn CM, van den Berg LH, Mane S, Mason CE, Choi M, Gaal E, Bayri Y, Kolb L, Arlier Z, Ravuri S, Ronkainen A, Tajima A, Laakso A, Hata A, Kasuya H, Koivisto T, Rinne J, Ohman J, Breteler MM, Wijmenga C, State MW, Rinkel GJ, Hernesniemi J, Jaaskelainen JE, Palotie A, Inoue I, Lifton RP, Gunel M. Susceptibility loci for intracranial aneurysm in European and Japanese populations. *Nat Genet*. 2008;40:1472–1477.
- Helgadottir A, Thorleifsson G, Magnusson KP, Gretarsdottir S, Steinthorsdottir V, Manolescu A, Jones GT, Rinkel GJ, Blankenstein JD, Ronkainen A, Jaaskelainen JE, Kyo Y, Lenk GM, Sakalihasan N, Kostulas K, Gottsater A, Flex A, Stefansson H, Hansen T, Andersen G, Weinsheimer S, Borch-Johnsen K, Jorgensen T, Shah SH, Quyyumi AA, Granger CB, Reilly MP, Austin H, Levey AI, Vaccarino V, Palsdottir E, Walters GB, Jonsson T, Snorrardottir S, Magnusdottir D, Gudmundsson G, Ferrell RE, Sveinbjornsdottir S, Hernesniemi J, Niemela M, Limet R, Andersen K, Sigurdsson G, Benediktsson R, Verhoeven EL, Teijink JA, Grobbee DE, Rader DJ, Collier DA, Pedersen O, Pola R, Hillert J, Lindblad B, Valdimarsson EM, Magnadottir HB, Wijmenga C, Tromp G, Baas AF, Ruigrok YM, van Rijn AM, Kuivaniemi H, Powell JT, Matthiasson SE, Gulcher JR, Thorgerisson G, Kong A, Thorsteinsdottir U, Stefansson K. The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nat Genet*. 2008;40:217–224.
- Pasman E, Laurendeau I, Heron D, Vidaud M, Vidaud D, Bieche I. Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res*. 2007;67:3963–3969.
- Wiebers DO, Whisnant JP, Huston J III, Meissner I, Brown RD Jr, Piegras DG, Forbes GS, Thielen K, Nichols D, O'Fallon WM, Peacock J, Jaeger L, Kassell NF, Kongable-Beckman GL, Torner JC. International Study of Unruptured Intracranial Aneurysms Investigators. Unruptured intracranial aneurysms: natural history, clinical outcome, and risks of surgical and endovascular treatment. *Lancet*. 2003;362:103–110.
- Wermer MJ, van der Schaaf IC, Algra A, Rinkel GJ. Risk of rupture of unruptured intracranial aneurysms in relation to patient and aneurysm characteristics: an updated meta-analysis. *Stroke*. 2007;38:1404–1410.
- Ruigrok YM, Rinkel GJ. Genetics of intracranial aneurysms. *Stroke*. 2008;39:1049–1055.
- Onda H, Kasuya H, Yoneyama T, Takakura K, Hori T, Takeda J, Nakajima T, Inoue I. Genomewide-linkage and haplotype-association studies map intracranial aneurysm to chromosome 7q11. *Am J Hum Genet*. 2001;69:804–819.
- McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007;316:1488–1491.
- Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, König IR, Stevens SE, Szymczak S, Tregouet DA, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, Tobin MD, Ziegler A, Thompson JR, Schunkert H. WTCCC and the Cardiogenics Consortium. Genomewide association analysis of coronary artery disease. *N Engl J Med*. 2007;357:443–453.
- Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447:661–678.
- Broadbent HM, Peden JF, Lorkowski S, Goel A, Ongen H, Green F, Clarke R, Collins R, Franzosi MG, Tognoni G, Seedorf U, Rust S, Eriksson P, Hamsten A, Farrall M, Watkins H. PROCARDIS consortium. Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNP in the ANRIL locus on chromosome 9p. *Hum Mol Genet*. 2008;17:806–814.
- Jarinova O, Stewart AF, Roberts R, Wells G, Lau P, Naing T, Buerki C, McLean BW, Cook RC, Parker JS, McPherson R. Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Arterioscler Thromb Vasc Biol*. 2009;29:1671–1677.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–575.
- Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet*. 2009;84:210–223.
- Begg CB, Zhang ZF. Statistical analysis of molecular epidemiology studies employing case-series. *Cancer Epidemiol Biomarkers Prev*. 1994;3:173–175.
- Deka R, Koller DL, Lai D, Indugula SR, Sun G, Woo D, Sauerbeck L, Moomaw CJ, Hornung R, Connolly ES, Anderson C, Rouleau G, Meissner I, Bailey-Wilson JE, Huston III J, Brown RD, Kleindorfer DO, Flaherty ML, Langefeld CD, Foroud T, Broderick JP. The FIA Study Investigators. The relationship between smoking and replicated sequence variants on chromosomes 8 and 9 with familial intracranial aneurysm. *Stroke*. 2010;41:1132–1137.
- Schunkert H, Gotz A, Braund P, McGinnis R, Tregouet DA, Mangino M, Linsel-Nitschke P, Cambien F, Hengstenberg C, Stark K, Blankenberg S, Tiret L, Ducimetiere P, Keniry A, Ghorri MJ, Schreiber S, El Mokhtari NE, Hall AS, Dixon RJ, Goodall AH, Liptau H, Pollard H, Schwarz DF, Hothorn LA, Wichmann HE, König IR, Fischer M, Meisinger C, Ouwehand W, Deloukas P, Thompson JR, Erdmann J, Ziegler A, Samani NJ. Cardiogenics Consortium. Repeated replication and a prospective meta-analysis of the association between chromosome 9p21.3 and coronary artery disease. *Circulation*. 2008;117:1675–1684.
- Leblanc R. Familial cerebral aneurysms. *Can J Neurol Sci*. 1997;24:191–199.
- Kasuya H, Onda H, Takeshita M, Hori T, Takakura K. Clinical features of intracranial aneurysms in siblings. *Neurosurgery*. 2000;46:1301–1305.
- Lindner SH, Bor AS, Rinkel GJ. Differences in risk factors according to the site of intracranial aneurysms. *J Neurol Neurosurg Psychiatry*. 2010;81:116–118.
- Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med*. 1992;116:78–84.
- Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA*. 1991;266:93–98.
- Mitchell P, Gholkar A, Vindlacheruvu RR, Mendelow AD. Unruptured intracranial aneurysms: benign curiosity or ticking bomb? *Lancet Neurol*. 2004;3:85–92.

日本臨牀 68 巻 増刊号 8 (2010 年 8 月 20 日発行) 別刷

遺伝子診療学(第2版)

—遺伝子診断の進歩とゲノム治療の展望—

I. 遺伝子診断(Genetic Diagnosis) A. 総論

遺伝子診療学研究の進歩

ゲノムインフォマティクス—システム医学に向けて—

中岡博史 井ノ上逸朗

I. 遺伝子診断 (Genetic Diagnosis) A. 総論

遺伝子診療学研究の進歩

ゲノムインフォマティクス—システム医学に向けて—

Genome informatics

中岡博史 井ノ上逸朗

Key words : 多因子疾患, genome-wide association study, phenome, システム医学

はじめに

近年, ヒトゲノム配列が解読され, 国際 HapMap プロジェクトによるヒト集団における遺伝的多型パターンの解明や一塩基多型 (single nucleotide polymorphism: SNP) のカタログ化が急速に進んでいる。同時に, 遺伝子型タイピング技術が飛躍的に向上したことで, ゲノム全域を網羅する 30 万–100 万の SNPs を用いた症例対照関連解析 (genome-wide association study: GWAS) が可能となり, 糖尿病や冠動脈疾患などの多因子疾患の感受性遺伝子が次々に同定されている¹⁾。2007 年には多くの感受性遺伝子の同定が Nature, Science 誌などで報告され, Science 誌の breakthrough of the year に human genetic variation が選ばれるなど注目を集めた。これらの成果は, 多因子疾患の発症メカニズムという複雑な生命現象の解明に向けて有用な情報となるであろう。しかし, GWAS で同定された SNP は疾患に対する遺伝的寄与のごくわずかしが説明していないことが指摘されている²⁾。現在の SNP タイピングプラットフォームではとらえきれない低頻度の多型 (rare variant, 集団内の頻度 < 5%) や構造多型 (structural variant) を検出し, ヒト集団における遺伝的多型を網羅的に収集し, 表現型との関連を精

査する必要があるだろう。また, 単一の疾患としての phenotype と genome の関連を評価するだけではなく, transcriptome, proteome, metabolome, diseasesome (疾患のサブタイプ, 疾患および疾患群) からなる多層的な総体である phenome と genome との関連に着目することが重要になると考えられる。多因子疾患の機構を理解し, 疾患の原因である生体システムの異常を特徴づける複合的表現型を把握することが, 疾患の診断や将来予測の精度向上につながる可能性がある。

本稿では, GWAS で得られた知見について概説するとともに, genome-phenome relationships 同定研究への展望について論じたい。

1. 多因子疾患の GWAS の現状と今後の展開

common disease-common variant 仮説は, ありふれた疾患の原因突然変異は共通の起源をもつというものである。この仮定の下, 全ゲノムを網羅する 30 万–100 万 SNPs を利用した症例対照関連解析を行うことで, 疾患関連座位を特定できる可能性がある。これは, 疾患に直接関連する変異とマーカー座位でのアレルの連鎖不平衡を利用するものである。現在, Affymetrix 社や Illumina 社が提供する全ゲノムをカバーす

Hirofumi Nakaoka, Ituro Inoue: Division of Molecular Life Science, School of Medicine, Tokai University 東海大学医学部 基礎医学系 分子生命科学

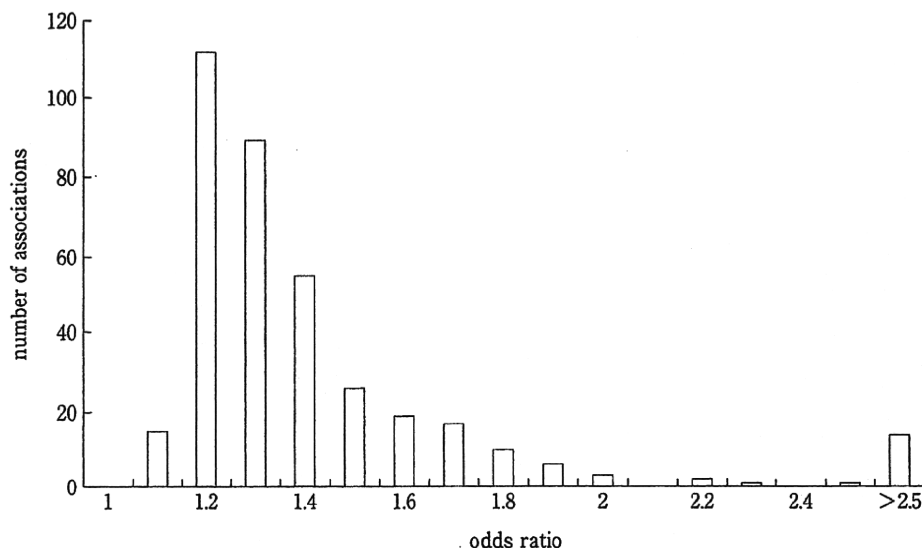


図1 多因子疾患のGWASで同定されたSNPのオッズ比に関する頻度分布
NHGRI GWA Catalog (<http://www.genome.gov/26525384>) のデータに基づき作成。

るタイピングプラットフォームを利用して、約80の多因子疾患および複雑形質において、数百に及ぶgenotype-phenotypeの関連が明らかになっている (<http://www.genome.gov/26525384>)。図1は、多因子疾患のGWASで検出されたSNPのオッズ比の頻度分布を示したものである。疾患のリスクを2倍以上増加させるSNPは少なく、ほとんどのSNPは1.2倍から1.4倍のリスク増加という弱い効果しかもっていないことがわかる。そのため、特定の疾患について同定されたすべての感受性SNPの効果を足し合わせても、家族集積性から推定される遺伝的寄与のごく一部しか説明できない。例えば、二型糖尿病に関する18の感受性遺伝子座で説明される遺伝的寄与の割合は6%にすぎない²⁾。このことは、当該疾患への遺伝的寄与が過大推定されている可能性を考慮しても、感受性遺伝子座位の大部分が未同定であることを示唆している。

現行のSNPタイピングプラットフォームは、全ゲノム上の5%以上の頻度を有するSNPの大部分をカバーするように設計されているものの、より低頻度な変異については十分にカバーしていない。rare variantはcommon variantより

も機能的である可能性が高いと考えられており³⁾、メンデル型遺伝病とGWASで特定されてきたSNPの中間的な効果の大きさをもつ低頻度の変異が疾患感受性と関連している可能性が指摘されている(図2)^{2,4)}。仮に、マイナーアレル頻度1%でオッズ比が3の多型を30座位同定することができれば、二型糖尿病の家族集積性を十分に説明できる⁴⁾。疾患関連解析におけるコピー数多型を含む構造多型の重要性も指摘されている⁵⁾。コピー数多型と疾患感受性との関連を調べる際には、その由来(*de novo*あるいは遺伝性)および頻度に着目する必要がある。Stefanssonらは、特定の領域に頻発する*de novo*な欠失が統合失調症と極めて強い関連を示すことを報告している⁶⁾。一方、集団内においてありふれたコピー数多型はSNPと連鎖不平衡にあるといわれており、疾患感受性に関連するコピー数多型を代理するSNPがGWASで検出されている可能性がある。実際、Crohn病のGWASで検出されたSNPと絶対連鎖不平衡($r^2=1$)にある20kbの欠失の存在が明らかにされている⁷⁾。

次世代シーケンサーを利用し、ヒト集団において1%以上の頻度を有する一塩基変異および

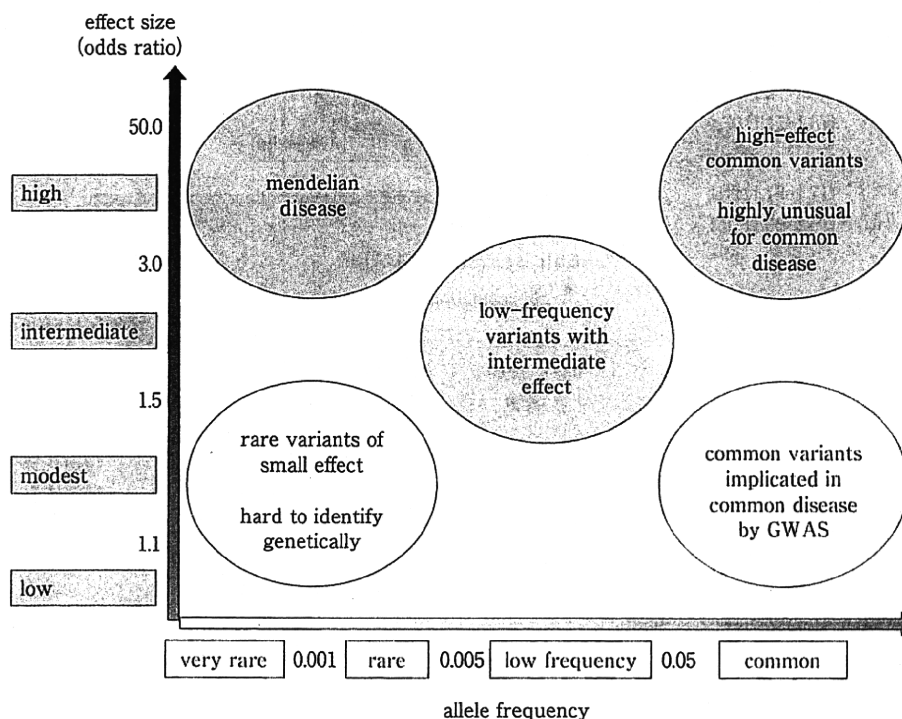


図2 アレル頻度と遺伝的効果のサイズ(オッズ比) (文献²⁴⁾より改変)

構造多型を網羅的に検出・カタログ化しようという試みである 1000 Genomes Project が 2008 年に立ち上げられた。これらの技術革新によってもたらされる膨大な情報により、ヒトゲノムに存在する変異の全体像をとらえ、ゲノムと疾患の関連を評価することが可能となるだろう。

2. Genome-phenome relationships

phenotype という用語は、生物の形態学的、生化学的、生理学的あるいは行動特性を表すものである。ゲノム医学では標準的な診断に基づいた疾患を表現型としてとらえている。病態が患者ごとに多様であり、幾つかの病型に分類される疾患の遺伝解析では、これら病型を個々の表現型(sub-phenotype)としてとらえ、関連解析を行うことで、特定の sub-phenotype と強く関連する感受性遺伝子を同定できる可能性がある。GWAS の創成期における目覚ましい成功例である加齢黄斑変性症(age-related macular degeneration: AMD)の GWAS では、明確な

表現型の定義に基づく症例・対照の選抜を行い、欧米人で主流であるドルーゼン形成が主体で脈絡膜新生血管がない萎縮型の dry AMD と呼ばれる病型と *CFH* 遺伝子に存在する非同義置換を起こす SNP が関連し⁸⁾、アジア人で頻度の高い脈絡膜新生血管を伴う滲出性の wet AMD の病型には *HTRA1* 遺伝子のプロモーター領域の SNP が関連すること⁹⁾を明らかにしている。

多因子疾患の GWAS で同定された SNP の多くはタンパク質配列に影響をもたらす変異ではないことから、これら SNP が近傍に存在する遺伝子の転写制御に関与しているのではないかと推測される。また、GWAS などにより同定された感受性領域が複数の疾患で一致する場合、これら一連の疾患は遺伝的に規定される分子生物学的ネットワークを共有すると考えられる。

その一例として、著者らが国際多施設共同研究による脳動脈瘤 GWAS¹⁰⁾によって同定した 9p21 領域を紹介する(図 3)。9p21 領域は複数の疾患の GWAS で関連が特定されている領域

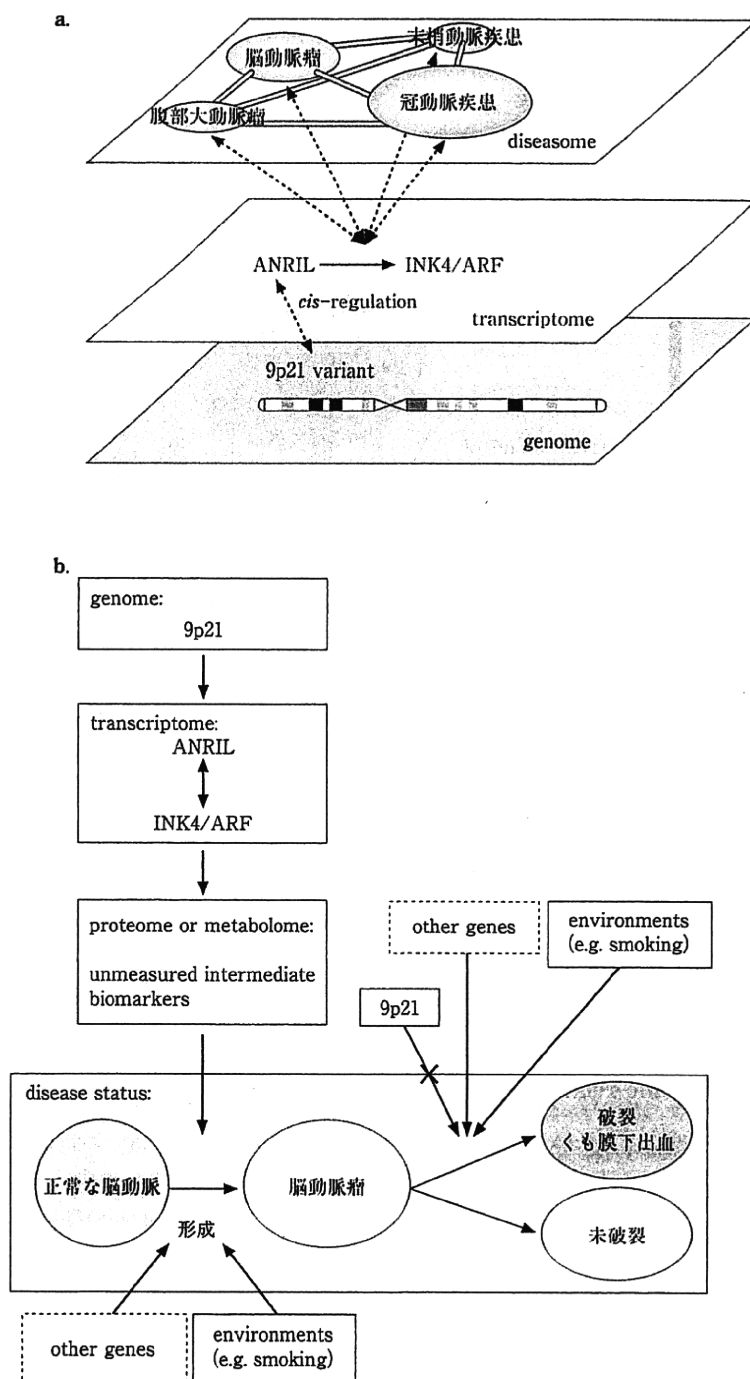


図3 脳動脈瘤感受性と9p21領域のSNPとをつなぐphenomeの多層的構造に関する概略図

a: 9p21領域の多型が共通のリスク因子である複数の動脈疾患(冠動脈疾患, 末梢動脈疾患, 腹部大動脈瘤そして脳動脈瘤)において推測される分子生物学的ネットワークの共有.

b: 脳動脈瘤のsub-phenotype解析の結果から推測される9p21領域のSNPと脳動脈瘤形成に関するネットワーク. 9p21領域のSNPは脳動脈瘤形成に関連するが, その破裂・未破裂の予測子とはならない. 脳動脈瘤形成や脳動脈の破裂のリスクには複数の遺伝子や環境の要因が関連している.

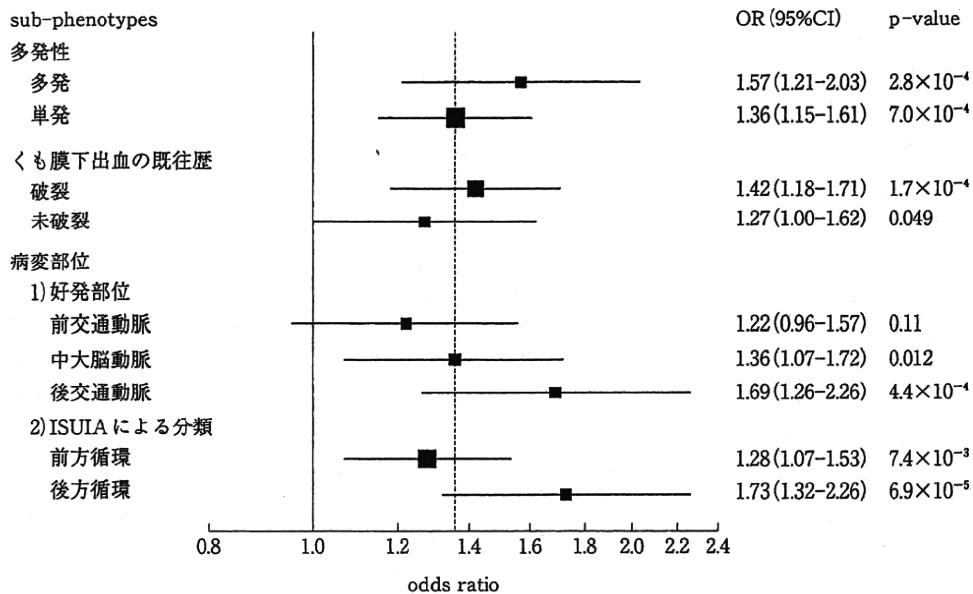


図4 9p21 領域の SNP rs1333040 の脳動脈瘤 sub-phenotypes に対するオッズ比
病変部位の分類(前方および後方循環)は International Study of Unruptured Intracranial Aneurysms (ISUIA)¹⁰⁾に基づく。

であり、9p21 領域に位置する SNP でタグされる配列多型が複数の動脈疾患(冠動脈疾患、末梢動脈疾患、腹部大動脈瘤そして脳動脈瘤)に関連することが明らかになっている¹¹⁾。これら動脈疾患の関連領域の近傍に存在するコード遺伝子は *CDKN2A* と *CDKN2B* である。*INK4/ARF* (*CDKN2A/B*) 領域は、3つの細胞周期阻害因子(*p16^{INK4a}*, *p15^{INK4b}* および *ARF*) をコードし、がん抑制遺伝子として働いている。しかし、冠動脈疾患や脳動脈瘤と強い関連を示した SNP は *INK4/ARF* 領域から約 100 kb 離れていた。9p21 感受性領域には長鎖の非コード RNA である *ANRIL* (NCBI の Entrez Gene データベースでは *CDKN2BAS* [*CDKN2B* antisense]) が存在している。*ANRIL* は血管内皮細胞、単球由来マクロファージや冠状動脈平滑筋細胞で発現することが確認されている。9p21 領域 SNP の遺伝子型と *ANRIL* の発現量が関連すること、同時に *ANRIL* の発現量が *INK4/ARF* の発現量とも関連していることが明らかにされた^{12,13)}。このことは、9p21 の感受性領域が分子スイッチとして働き、*ANRIL* の転写量を調節することで、細

胞周期制御遺伝子の発現量が変化するという、細胞増殖に関する分子ネットワークに変動が生じ、血管リモデリングなどの異常を介して、これら動脈系疾患の発症に至るのではないかと推測されている(図3-a)。

脳動脈瘤は破裂・未破裂、多発性、病変部位や瘤の大きさなど様々な sub-phenotype をもち、これら sub-phenotype が患者の転帰と関連している。そこで、著者らが脳動脈瘤感受性遺伝子座として同定した 9p21 領域の SNP (rs1333040) が、特定の sub-phenotype に特異的な効果を示す可能性について評価した。図4にサブグループ特異的オッズ比の推定結果を示す。脳動脈瘤の破裂・未破裂のサブグループ間および脳動脈瘤の多発・単発の間には有意なオッズ比の差異は認められなかった。病変部位によるサブグループ解析の結果、好発部位の中で、後交通動脈に脳動脈瘤をもつサブグループで最も強い関連が認められた。一方、前交通動脈の脳動脈瘤と rs1333040 の関連は有意ではなかった。更に、病変部位を前方循環と後方循環に分類したサブグループ解析を行ったところ、rs1333040 と前

方循環の脳動脈瘤との関連は弱く、後方循環の脳動脈瘤に対して強い関連を示すことがわかった。前方循環と後方循環の脳動脈瘤に対するオッズ比の差異は有意なものであった($p=0.03$)。これらの結果から、9p21 領域の SNP は脳動脈瘤の病変部位に関連していることが示唆され、特に後方循環の脳動脈瘤形成に関与していることが明らかになった(Nakaoka ら、投稿中)。脳動脈瘤の破裂・未破裂のサブグループ間に差異が認められなかったことから、9p21 の感受性領域は脳動脈瘤の成長および破裂に影響を及ぼす因子ではなく、瘤の形成段階に関与していると推察される。図 3-b でモデル化された脳動脈瘤の形成パスウェイの中で、proteome および metabolome の階層に発露する特性は現在のところ同定されていない。この多層的関連構造をつなぐ新規な生化学的バイオマーカーを発見できれば、非常に有用な疾患発症の予測子となるだろう。また、脳動脈瘤の sub-phenotype で最も重要な将来の破裂のリスクと関連する遺伝的予測子の同定が求められる。

ここまでは、特定の表現型(疾患)に注目して、GWAS により感受性遺伝子を同定し、その遺伝子変異により対象とする表現型が生じるメカニズムを解明するという、順遺伝学的研究について紹介してきた。近年では、遺伝子発現量や代謝産物量についてハイスループット測定技術が確立されてきたため、遺伝子発現量や代謝産物量の変動に関与する遺伝的多型を特定し、更にその遺伝的多型がどのような表現型(疾患)と関連しているかについて評価することで、疾患の感受性遺伝子やその背景となるメカニズムを同時に特定しようという、逆遺伝学的な試みも行われるようになった。Gieger ら¹⁴⁾は、血清中のメタボライト(代謝産物)濃度を網羅的にプロファイリングし、それらの変動に関連する SNP を GWAS により特定した。次いで、メタボライ

ト濃度と関連する SNP が、HDL コレステロールなどの臨床的指標、更には冠動脈疾患などの臨床的エンドポイントとも関連することを示した。これらは、メタボライトを疾患の中間表現型(intermediate phenotype)として用いることで、疾患のサブグループ化や複数の疾患で共通する発症メカニズムの解明に向けて新たな方向性を創出できることを示唆している。また、メタボライトの測定値の変動に対する SNP の寄与率は非常に高いことも明らかにしている(1 SNP の表現型分散への寄与率が 12% にも及ぶ)。今後、個人が曝露されている環境ごとにメタボローム・プロファイルを得ることで、疾患を引き起こすメカニズムにおける遺伝と環境の交互作用の解明も可能となり、個の医療の実現につながると期待できる。

phenome を系統的に研究するためには、疾患への深い知識、phenotype の網羅的な収集を実現するプラットフォームの開発、更に多因子疾患の GWAS で培われた解析手法を基礎として phenome の多層的構造を考慮に入れた数理モデルによる解析が必要になると考えられる。同一検体から得られた幾つかの phenotype 間に、階層構造から推論される方向性をもたせ、因果モデルを構築する手法としては、構造方程式モデルを用いたパス解析のような手法が有効となるだろう¹⁵⁾。

phenomics には医学、遺伝学、分子生物学、細胞生物学、システム生物学、統計学、情報学分野などの融合による超学際的な‘システム医学’という分野を開拓する必要があるだろう。また、単独の研究機関が収集できる phenome は限定的なものであることから、研究目的に応じて収集する phenotype の選別、あるいは phenome データベース間の情報共有も視野に入れて研究を進める必要があるだろう。

■ 文 献

- 1) Manolio TA, et al: A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* 118: 1590-1605, 2008.
- 2) Manolio TA, et al: Finding the missing heritability of complex diseases. *Nature* 461: 747-753, 2009.
- 3) Gorlov IP, et al: Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am J Hum Genet* 82: 100-112, 2008.
- 4) McCarthy MI, et al: Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9: 356-369, 2008.
- 5) McCarroll SA: Extending genome-wide association studies to copy-number variation. *Hum Mol Genet* 17: R135-142, 2008.
- 6) Stefansson H, et al: Large recurrent microdeletions associated with schizophrenia. *Nature* 455: 232-236, 2008.
- 7) McCarroll SA, et al: Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nat Genet* 40: 1107-1112, 2008.
- 8) Klein RJ, et al: Complement factor H polymorphism in age-related macular degeneration. *Science* 308: 385-389, 2005.
- 9) Dewan A, et al: HTRA1 promoter polymorphism in wet age-related macular degeneration. *Science* 314: 989-992, 2006.
- 10) Bilguvar K, et al: Susceptibility loci for intracranial aneurysm in European and Japanese populations. *Nat Genet* 40: 1472-1477, 2008.
- 11) Helgadóttir A, et al: The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nat Genet* 40: 217-224, 2008.
- 12) Liu Y, et al: INK4/ARF transcript expression is associated with chromosome 9p21 variants linked to atherosclerosis. *PLoS One* 4: e5027, 2009.
- 13) Jarinova O, et al: Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Arterioscler Thromb Vasc Biol* 29: 1671-1677, 2009.
- 14) Gieger C, et al: Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet* 4: e1000282, 2008.
- 15) Greenland S, Brumback B: An overview of relations among causal modelling methods. *Int J Epidemiol* 31: 1030-1037, 2002.
- 16) Wiebers DO, et al: International Study of Unruptured Intracranial Aneurysms Investigators. Unruptured intracranial aneurysms: natural history, clinical outcome, and risks of surgical and endovascular treatment. *Lancet* 362: 103-110, 2003.

表 題

著 者 名

週 刊
医学のあゆみ 別 刷

第 卷・第 号： 年 月 日号

ゲノム情報を含む医学情報処理の将来

The future of information processing in medical genomics



中岡博史(写真) 井ノ上逸朗

Hirofumi NAKAOKA and Ituro INOUE

国立遺伝学研究所総合遺伝研究系人類遺伝研究部門

◎近年、ゲノム全域関連解析により多因子疾患の感受性遺伝子の同定が行われているが、臨床応用をめざすうえで重要となるのは疾患の特性である生体システムの異常を特徴づける複合的な表現型の詳細な情報取得であり、デジタル化である。大規模な臨床情報のデジタル化に伴う情報蓄積を行い、個々人のゲノム情報というデジタル信号と結合する、医学・医療におけるデータ集約型サイエンス/ヘルスケアの実現が求められる。このことにより疾患バイオマーカーの発見、標的治療法の開発、個の医療の実現、ゲノム情報による創薬といった、当初ヒトゲノム計画が設定したゴールに近づくこととなる。本稿では、今後の医学における情報解析の方向性および疾患メカニズムのシステム解析における新しい展開について論じる。



Key word : ゲノム全域関連解析, パーソナルゲノム情報, データ集約型サイエンス/ヘルスケア, ネットワーク解析

近年、ゲノム全域関連解析(genome-wide association study: GWAS)により、糖尿病や冠動脈疾患などの多因子疾患の感受性遺伝子がつぎつぎに同定されている¹⁾。2007年には多くの感受性遺伝子の同定が『Nature』、『Science』誌などで報告され、『Science』誌の breakthrough of the year に“Human Genetic Variation”が選ばれるなど注目を集めた。これらの成果は、多因子疾患の発症メカニズムという複雑な生命現象の解明に直結すると期待されたものの、ひとつひとつの遺伝子多型の疾患への関与が非常に小さいこともあり、疾患との直接的な機能的関連は不明である。統合的疾患理解には革新的解析手法の開発が望まれる。

ゲノム医科学は、ヒトのもつ表現型の多様性とDNA配列多型との関連を追及するという研究背景から、そこで扱われるDNA配列のもつ膨大な情報量に注目が集まった。近い将来には1人につき1,000ドルでの全ゲノム配列決定が実現すると予想され、二倍体のヒトゲノム約60億塩基対における個々人のDNA多型情報が得られることになる。臨床応用をめざすうえで重要となるのは、疾患の特性である生体システムの異常を特徴づける

複合的な表現型の詳細な情報取得であり、デジタル化である。かつ個々の表現型とゲノムの関連だけでなく、疾患ネットワークとしての理解が求められる。そこでは分子表現型(transcriptome, proteome, metabolome)およびdiseasome(疾患のサブタイプ、疾患および疾患群、併存疾患の情報)といった表現型の総体(phenome)とパーソナルゲノム情報との関連に着目することが重要になると考えられる。

データ集約型サイエンス/ヘルスケア

2003年4月、ヒトゲノム完全解読の報告がなされた。これは生命現象を担うデジタル情報をわれわれがはじめて手にいれ、生命科学に転換が起こったときとして歴史に記録されることとなる。さらにはこの数年、個人のゲノム情報に注目が集まり、テクノロジーの進歩に従い安価・迅速に個人の全ゲノム情報を得ることが可能となってきた。5~10年後には、医療の現場で個人ゲノム情報が診断・治療方針の決定に利用されるであろう。2009年、アメリカ合衆国ではオバマ大統領による景気対策の一環でもあろうが、電子カルテの

開発・普及に年間2兆円、5年にわたり支援することが決まった。国家が電子カルテ化を支援することにより個々の医療施設の診療上の利便性のためではなく、アメリカ中の施設においてデジタル化された医療情報の蓄積・集約および共有が可能となる。すでに Medicare に登録されている1,300万人規模の臨床情報を用いた疾患表現型ネットワーク解析がなされている²⁾。当然、今後、医療情報は膨大なものとなることが予想され、ヒト固有のデジタル情報(ゲノム)と結合することにより、さまざまな中間表現型を含む疾患のデジタル的理解、そして革新的治療法開発へとつながる。また、そのための統計的解析手法開発も研究対象となる。医学・医療において“情報”が革新的解析法開発の鍵となり、悲観的な見方では情報洪水といわれる状況ともなりうるが、“情報”を駆使した知識発見(information driven knowledge)が重要であろう。

マイクロソフト研究所の Jim Gray は 4th paradigm という概念を提唱している。経験型サイエンスではじまり、理論サイエンス、コンピュータサイエンスと進展してきたが、第4番目のパラダイムとしてデータ集約型サイエンスが位置づけられると提唱している。ゲノムサイエンスの進展もあり、サイエンスに供せられるデータの規模は今後さらに増大していくであろう。そこでデータをどのように活用し新知見を得るかが重要であり、data curation あるいは因子化(因数分解による機能的因子の同定)といった知識獲得手法の開発が鍵となる。医学・医療において、他の生命科学より先んじてデータ集約型ヘルスケアとして第4のパラダイムが実現される可能性がある。それは前述のように数千万人規模での臨床情報のデジタル化に伴う情報蓄積のみならず、個々人のゲノム情報というデジタル信号との結合が可能であるからである。社会の要請も高まるに違いない。かつゲノム情報を結合したデータ集約型ヘルスケアは、つねに個人を対象にしていることも注目すべきである。これにより疾患バイオマーカーの発見、標的治療法の開発、個の医療の実現、ゲノム情報による創薬といった、当初ヒトゲノム計画が設定したゴールに近づくこととなろう。以下ではゲノム

情報と表現型を結合するネットワーク解析に重点をおき、疾患メカニズムのシステム解析における新しい展開および今後の医学における情報解析の方向性を考える。

疾患表現型の詳細な記述による分離と統合

ゲノム医学では、標準的な診断に基づいた疾患を表現型としてとらえている。病態が患者ごとに多様であり、いくつかの病型に分類される疾患の遺伝解析ではこれら病型を個々の表現型(sub-phenotype)としてとらえ関連解析を行うことで、特定の sub-phenotype と強く関連する感受性遺伝子を同定できる可能性がある。GWAS の創成期におけるめざましい成功例である加齢黄斑変性症(age-related macular degeneration: AMD)の GWAS では、明確な表現型の定義に基づく症例・対照の選抜を行い、欧米人で主流であるドルーゼン形成が主体で脈絡膜新生血管がない萎縮型の dry AMD とよばれる病型と *CFH* 遺伝子に存在する非同義置換を起こす SNP が関連し³⁾、アジア人で頻度の高い脈絡膜新生血管を伴う滲出性の wet AMD の病型には *HTRA1* 遺伝子のプロモーター領域の SNP が関連すること⁴⁾を明らかにしている。著者らは国際多施設共同 GWAS⁵⁾で脳動脈瘤感受性遺伝子座として同定した 9p21 領域の SNP (rs1333040)が、脳動脈瘤のサブ表現型(破裂・未破裂、多発性、病変部位や瘤の大きさ)に特異的な効果を示す可能性について評価し、9p21 の感受性領域が脳動脈瘤の成長および破裂に影響を及ぼす因子ではなく、とくに後方循環の瘤の形成に強く関連していることを明らかにした⁶⁾。

一方、複数の疾患が同一の遺伝子と関連していることが知られており、共通のメカニズムにおける異常が複数の疾患を引き起こすと考えられる⁷⁾。つまり、疾患はたがいに排他な事象ではなく、たがいにオーバーラップする部分があると考えられることもできる。そこで、単一の疾患を“表現型”として扱うだけではなく、患者の病歴のように、“たえず更新される表現型”から得られる疾患の共存関係(comorbidity)を精査することで疾患の総体の理解を進めようという試みがなされている。Hidalgo らは、アメリカ政府の医療保険(Medicare)

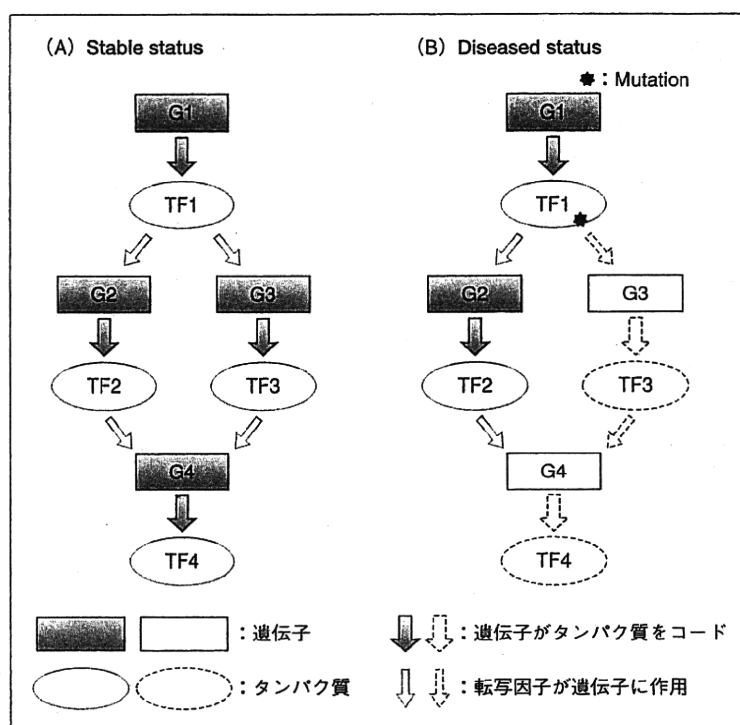


図 1 生体システムの異常としての疾患

A: 安定状態, B: 病的状態.

長方形は遺伝子, 楕円はタンパク質を表す. 太い矢印は遺伝子がタンパク質をコードしていることを, 細い矢印は転写因子が遺伝子に作用することを示す. 病的状態では遺伝子(G1)の変異が引き起こす転写因子(TF1)の構造変化により遺伝子発現ネットワークに異常をきたし, 結果として疾患の発症へとつながることを示している.

に登録された 1,300 万人, 3,200 万件の保険料請求記録に基づき, 同一患者の併存疾患の情報を精査し, 疾患表現型間のネットワークを構築している. そのネットワーク力学的解析により, 患者の病歴の推移はネットワーク内で近接する疾患へと移行する可能性が高いこと, ネットワーク内で他の多くの疾患と強いつながりをもつ疾患は患者の病歴における後期の疾患である可能性が高く, かつ致死性も高いものであることを明らかにしている²⁾. さらに Park らは, Hidalgo らが構築した疾患表現型ネットワークにおける併存疾患の背景にある分子メカニズムを, 大規模なデータベース検索に基づいて同定している⁸⁾.

生体システムの異常としての疾患

近年, 疾患を生体システムのネットワークの異常としてとらえ, 当該疾患のネットワークにかか

わる分子の挙動や制御について解析することで, 複雑な疾患をより精密に理解しようという先進的な試みがなされている^{9,10)}. 定常的な状態にある生体システムに, 小さな変動(摂動; perturbation), たとえば遺伝的な変異や非遺伝的な環境の変化が与えられたとき, 生体システムは定常的な状態からすこしだけ変化すると考えると, 疾患状態とは何らかの摂動によって生体システムに大きな崩れが生じた状態であると考えられる. 生体システムは多様な摂動に対して頑健性をもつことでその機能を維持しているが, 特定の摂動に対してはきわめて脆弱である場合があると考えられている. たとえば, ヒトゲノムに存在する変異の大多数はある特定の表現型に何ら影響を及ぼさないが, 特定の遺伝子の変異が原因となり遺伝病を引き起こすことが知られている(Huntington 病, 嚢胞性線維症など). 図 1 は遺伝子 1(G1)の変異が引き起こす

転写因子 1(TF1)の構造変化により遺伝子 3(G3)への結合が損なわれ、結果として遺伝子発現ネットワークが正常に機能しなくなることによって疾患の発症につながるという模式図である。一般に、図 1 に示すような疾患ネットワークは既知ではないため、網羅的な分子表現型(transcriptome, proteome, metabolome)の測定を行い、病的状態を分子表現型による複合的表現型として特徴づける必要がある。図 1 の場合、遺伝子 3, 4(G3, 4)の転写レベルによってシステムの異常をとらえられる可能性がある。

疾患の状態によって変化する生体ネットワークを同定するためには、データベース化された分子表現型間の相互作用ネットワークの情報を利用することが有効である。Taylor らはヒトのタンパク質間相互作用ネットワークにおいて、多数のタンパク質と相互作用関係にあるハブタンパク質(hub protein)が存在することに着目した。彼らは癌組織での全ゲノム遺伝子発現解析によって、ハブタンパクをコードする遺伝子のうち 256 遺伝子で、乳癌の予後に応じて相互作用パートナーとの共発現構造が顕著に変化していることを示している¹¹⁾。

遺伝子発現ネットワークを用いた GWAS の精緻化

前述のように多因子疾患で GWAS が行われ、多数の感受性遺伝子同定が報告されている。ここで注意が必要なのは多くの場合、疾患との関連が示された DNA 多型の近傍に存在する遺伝子を便宜的に“感受性遺伝子”とよんでいるだけであり、その遺伝子が実際に疾患の発症に関与しているという証拠が得られていない場合が多数であるという点である。GWAS で同定された疾患と関連を示す SNPs のうち、80% 以上が遺伝子間領域かイントロンに存在する SNPs であり、アミノ酸配列に変化を起こす非同義置換の SNPs は 10% に満たない¹²⁾。このことは GWAS で同定された DNA 多型の多くは、近傍に存在する遺伝子の転写制御に関与している可能性が高いことを示している¹³⁾。そこで、疾患と関連する DNA 多型に応じて変化する分子ネットワーク(causal network)を理解す

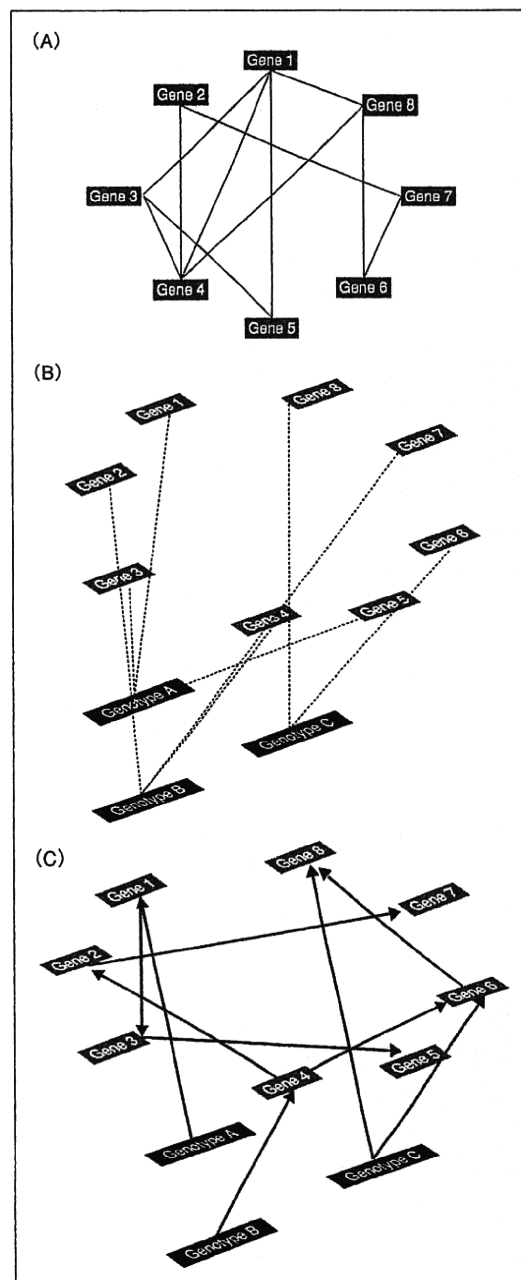


図 2 遺伝子共発現構造およびeQTL情報の統合による遺伝子発現ネットワークの構築

A: 遺伝子の共発現構造。実線でつながっている遺伝子間には発現量の相関関係が存在することを示す。

B: 遺伝子発現量と関連する遺伝子型(eQTL)の情報。各遺伝子座 A~C における遺伝子型が破線でつながっている遺伝子の発現量と関連していることを示す。

C: 遺伝子型多型によって規定される遺伝子発現ネットワーク。A および B の情報を統合し、遺伝子発現の因果関係を明確にしている。例: 遺伝子型 A が遺伝子 1 の発現量に変化を生じさせ、それに応じて遺伝子 3、ついで遺伝子 5 の発現量が変化する。

ることで、GWASの知見を新規バイオマーカーの創出や創薬へと繋げることが可能になると考えられる。

疾患の causal network の理解には、当該疾患と関係のある組織における分子表現型(transcriptome, proteome, metabolome)の共発現の情報や分子表現型とDNA多型との関連を精査し、疾患とDNA多型との関連の情報と統合することが必要となる(図2)。ある特定の組織で網羅的遺伝子発現量の計測を行うことで、発現量がたがいに相関する遺伝子群の情報、つまり遺伝子の共発現構造が明らかになる(図2-A)。さらに、遺伝子発現量と関連するDNA多型(expression quantitative trait loci: eQTL)を同定し(図2-B)、両者の情報を統合的に解析することで、DNA配列多型によって規定される遺伝子発現ネットワークを明らかにすることができる(図2-C)。このようにDNA多型が規定する遺伝子発現ネットワークの情報を構築することによって疾患とDNA多型との関連が同定された場合、当該DNA多型によって変動する遺伝子発現ネットワークの推測が可能となり、そのネットワーク上に存在する遺伝子を疾患の感受性遺伝子の候補として同定することが可能になる。

DNA配列多型によって規定される whole-gene expression ネットワークを構築するためには、gene expression traits 間の依存関係を説明する最適モデルを選択するための統計情報学的手法が必要となる。

本稿では、ベイジアンネットワークを用いた解析手法を紹介したい。ベイジアンネットワークは、図2-Cのように矢印をたどっても元に戻らない有向非巡回グラフとよばれるグラフィカルな表現に基づき、因果推論を確率モデルとして扱う統計手法である。ベイジアンネットワークの利点はきわめて複雑な系に対しても、ノード間の条件付き確率を用いて依存関係を確率のモデルとして比較的容易に扱うことができる点である。ベイジアンネットワークを考えるうえで重要な概念のひとつである“マルコフ性”について図3で説明する。

実際に whole-gene expression traits から構成されるグラフについて考える場合、ほぼ無数に存在するグラフのなかから gene expression traits 間の

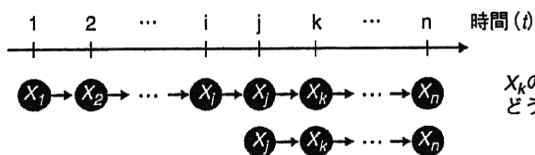
依存関係をよりよく説明する最適モデルを選択する必要がある。Schadtらは一連の研究を通じて、それぞれの gene expression trait の変動に関連しているDNA多型(eQTL)の情報を取り込むことで、発現量が相関関係にある複数の遺伝子間の因果関係(どの遺伝子が causal driver か)を推測する手法を提示している¹⁴⁻¹⁸⁾。それは“もし任意の2つの遺伝子の発現量が強く相関しており、それら遺伝子発現がDNA多型によってコントロールされている場合、両者のeQTLの多くはオーバーラップするであろう”という仮説に基づいている。

図4-Aのように、遺伝子発現データだけを用いた場合、遺伝子XとYの発現量に相関関係が存在するということはわかるが、どちらが causal driver か、つまりどちらが親ノードかという情報は得られない。そこで、発現量が相関関係にある複数の遺伝子間の因果関係(どの遺伝子が causal driver か)をeQTLの情報に基づいて推測するための系統的なルールとして、遺伝子間の依存関係に対する“主観確率(事前情報)”を導入している^{14,16,17)}。ここで、“主観確率”とは統計家(つまり解析を行う本人)が、ある事象が生起することに対してどれほどの確信度(degree of belief)をもっていかを定量的に表現するための概念である。遺伝子間の依存関係に対する主観確率の一例を図4-B,Cに示す。

“whole gene-expression array で観測された遺伝子発現量のデータ”と“eQTLの情報に基づくノード間の依存関係に対する主観確率”を結びつけて統計学的推測を行うために、ベイズ統計学が有用である。図4-Dはベイズ統計の推論過程を簡単に図示したものである。まず、事前の情報(ネットワークモデルに対する統計家の主観的な確率)があり、データを観測するという経験を通して事前の情報が更新され、ネットワークモデルに対する事後情報が得られる。

上記のようにeQTLの情報に基づくノード間の依存関係に対する主観確率をベイジアンネットワークに取り込むことは、考慮すべきネットワークの数に制限を加え計算負荷を抑えることだけでなく、現在の生物学的知見に照らし合わせて“もっともらしい”と考えられる一定の方向付けを可能

マルコフ過程：確率変数 X の過去への依存関係が直前のみへの依存



X_k の生起の仕方は、直前の X_j さえ決まっていれば、 i 番目以前がどうであったかには無関係

$$p(X_k | X_i, X_j, \dots, X_l) = p(X_k | X_i) \quad (1)$$

X_5 の祖先

x_5 の親ノード

$$p(X_5 | X_1, X_2, X_3, X_4) = p(X_5 | X_3, X_4)$$

子ノードの生起の仕方は親ノードのみに依存し、
それ以前の祖先ノードには依存しない

$$X=\{X_1, X_2, \dots, X_n\}$$
からなるDAGの同時確率

$$p(X_1, X_2, \dots, X_n) = \prod_{i=1}^n p(X_i | pa(i)) \quad (2)$$

$pa(i):X_i$ の親の集合

図 3 ベイジアンネットワークにおけるマルコフ性

A: マルコフ性を満たす確率過程. 確率変数 $X_1, X_2, X_3, \dots$ が 1, 2, 3...の時間の順に並んでいて, $k > j > i > \dots$ について, 確率変数 X の過去への依存関係が直前のみに依存する, つまり X_k の生起の仕方は, 直前の X_j さえ決まっていれば i 番目以前がどうであったかには無関係であること (マルコフ性) を満たす確率過程をマルコフ過程という. 確率表現は図中①の式となる.

B: 複雑なグラフにおけるマルコフ性. 遺伝子発現ネットワークのように複雑な系ではネットワークは広がりを持ち、各ノードは複数の子ノードをもち、また同時に複数の親ノードをもち、その親ノードにも親ノードが存在する。その場合、有向非巡回グラフにおけるマルコフ性は、子ノードの生起の仕方は(直接エッジでつながっている)親ノードのみに依存し、それ以前の祖先ノードには依存しないことと定義される。ノード X_i の親の集合を $pa(i)$ とすると、一連のノード $X = \{X_1, X_2, \dots, X_n\}$ からなる有向非巡回グラフは図中②の同時確率である。

にしている。そのため構築したネットワークは、DNA 多型によって生じる遺伝子発現のシステムティックな変化を反映したものになると考えられる。そのネットワークから得られる情報は、GWAS

の結果の理解に有用なものとなりうる。たとえば、ある疾患の GWAS で同定された DNA 多型により、どの遺伝子の発現量が変化し、その多型に応じてともに発現量が変化する一連の遺伝子群の情

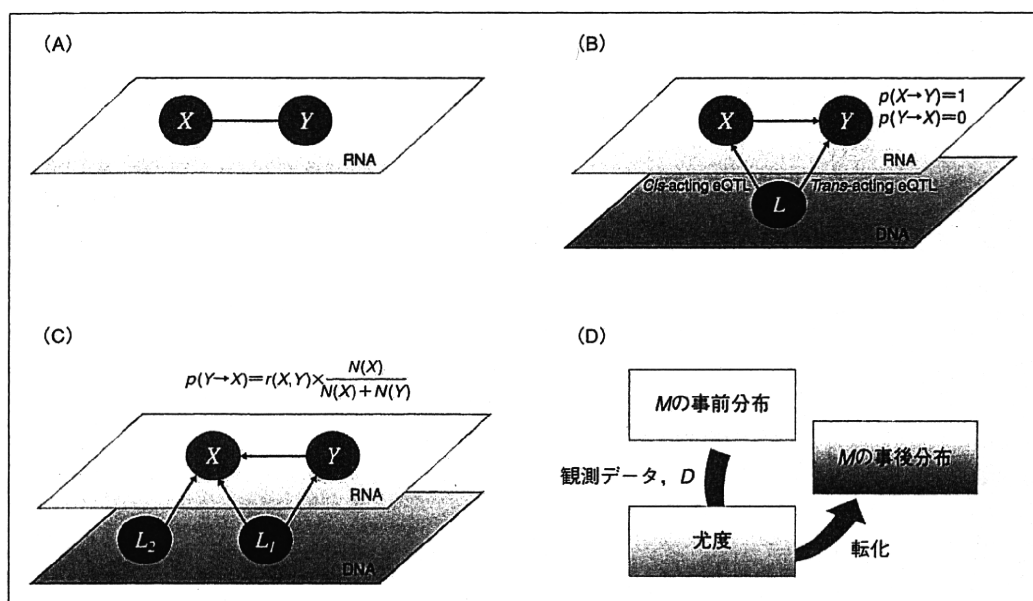


図 4 遺伝子発現ネットワーク構築における遺伝子間の因果関係を推定するための主観確率(事前情報)の導入

A: 遺伝子 X と Y の発現量が相関関係。

B: 同一の eQTL をもつ遺伝子発現の因果関係に対する主観確率。X が L の近傍に存在する cis-acting eQTL であり、Y が L と物理的に離れて存在する trans-acting eQTL である場合、causal network として eQTL、L が分子スイッチとして働いて X の発現量を制御し、その変化に応じて Y の発現量が変化するというモデルが自然であると考えられるため、Zhu ら (2004) は、この場合の遺伝子 X、Y の依存関係に対する主観確率を $\{p(X \rightarrow Y) = 1, p(Y \rightarrow X) = 0\}$ と定めている。

C: 同一の eQTL をもつ遺伝子発現の因果関係に対する主観確率。遺伝子 X、Y が共通して cis-acting eQTL (あるいは trans-acting eQTL) をもっている場合に、遺伝子 Y は eQTL として L_1 のみをもち、遺伝子 X は eQTL として L_1 と L_2 をもっていると仮定すると、もし X が Y を制御しているのであれば、X の eQTL である L_2 も Y の eQTL となると推測されるため、Y が X に対する親ノードとなる可能性が高いと考えられる。Zhu ら (2004) は、この場合の遺伝子 X、Y の依存関係に対する主観確率を

$$\begin{cases} p(X \rightarrow Y) = r(X, Y) \times \frac{N(Y)}{N(X) + N(Y)} \\ p(Y \rightarrow X) = r(X, Y) \times \frac{N(X)}{N(X) + N(Y)} \end{cases}$$

と定めている。 $r(X, Y)$ は X と Y の相関係数、 $N(\)$ は各遺伝子の eQTL の数を表す。上式は、より少数の eQTL をもつ遺伝子が親ノードになりやすいという主観を取り入れている。

D: ベイズ統計の推論過程。Whole gene expression array で全転写産物の発現量の測定値 D を得たとき、その原因となっている遺伝子発現ネットワークが M という有向非巡回グラフである確率の推論。

報が得られ、その遺伝子群に多く認められる分子機能や生物学的プロセスなどをデータベースで検索することで、疾患の causal pathway を推測するのに役立つと考えられる。

実際に Schadt らは上記の手法を用いて、ヒト肝組織における DNA 配列多型によって規定される遺伝子発現ネットワークを構築し、先行研究において構築したマウス肝組織でのネットワークの情報を併用して、近年報告された糖尿病や冠動脈疾患における GWAS の結果の理解を精緻化するこ

とを試み、GWAS の原著^{19,20)}で“感受性遺伝子”として報告されていた遺伝子とは別の遺伝子が当該疾患の原因に深く関与しているという推論を導き出している¹⁵⁾。

おわりに

2010 年 10 月に 1,000 genomes project の pilot phase の結果が公開され、低頻度な多型をも網羅するヒト集団における遺伝的多様性のパネルが利用可能になった²¹⁾。これらヒト多様性の情報によ