

環境因子と遺伝子の相互作用

Gene-Environment Interaction

病気は遺伝的素因と環境因子で決定されます。よって病因を解明するためには遺伝子と環境の相互作用に関する研究を避けて通ることはできません。殊に、Human Genome Project の成果により従来の家系図などを中心に理論展開してきた Genetic epidemiology も大きな変貌を遂げつつあります。本章では遺伝的素因と環境因子をどのように結びつけるかについての臨床研究の手法について解説します。

Paraoxonase と疾患

近年遺伝子の多様性が素因、すなわち病気のなり易さ(susceptibility)に結びつくことがわかってきました。遺伝子変異は一定の確率で発生しますが、その変異が致命的でなければ子孫に伝えられていきます。そしてある地域、ある環境において浸透度が一定します。例えば paraoxonase は農薬の解毒酵素ですが、アミノ酸 192 番目に変異をもち酵素活性は Gln/Gln < Gln/Arg < Arg/Arg の順で 10 倍以上異なります。仮にこの農薬の解毒酵素が Gln/Gln で低い家系は流産しやすかったり、精子の数が減少し不妊症になりやすいとしますと、人口分布は徐々に Arg/Arg の割合が増えることになります。この Gln/Gln は白人で多く、アジア人で少ないことが判明しました。paraoxonase は HDL に結合し、LDL による脂肪酸化を防ぐことにより動脈硬化を防ぐことが最近示されました。冠動脈疾患は paraoxonase 活性の低い Gln/Gln で多いとする報告もあります。日本では悪性腫瘍が死因のトップですが、アメリカでは心臓病であり、paraoxonase 活性がアメリカ人に低い傾向があることに心臓病が多い理由を求めることもできるかもしれません。

遺伝子 環境相互作用(gene-environment interaction)

まずは遺伝的素因を規定する因子が 1 つであり、環境中に存在する危険因子も 1 つで暴露されたかされなかったか(1 or 0)で決定されるような単純なものから考えていきましょう。その病気にかかりやすい遺伝的素因もなく環境中に存在する危険因子に暴露されていない人の一定の確率で病気になると仮定し、そのリスクを 1 とします。そうすると病気に対する危険環境因子には暴露されなかったが危険遺伝的素因を持つものは IR_g 、環境因子に暴露されたが遺伝的素因を持たないものは IR_e 、遺伝的素因をもち環境因子にも暴露された人は IR_{ge} のリスクを持つことになります。さらにそれぞれを 1 で割ると遺伝的素因による、環境因子暴露による、そして両者による相対比を計算することができます。すなわち、 R_e は遺伝的素因はないが危険環境因子に暴露された人が病気になるリスクを、遺伝子素因もなく危険環境因子にも暴露していない人の病気になるリスクで割ったもの、 R_g は 遺伝的素因を持つが危険環境因子に暴露されていない人が病気になるリスクを、遺伝子素因もなく危険環境因子にも暴露していない人の病気になるリスクで割ったもの、 R_{ge} は 遺伝的素因を持ち危険環境因子にも暴露された人が病気になるリスクを、遺伝子素因もなく危険環境因子にも暴露していない人の病気になるリスクで割ったものです。特に R_{ge} は遺伝子 環境相互作用の強さをあらわしています。

Cohort study				Case control study		
環境因子暴露	遺伝的素因	病気リスク	RR	cases	Control	OR
0	0	I	1	A ₀₀	B ₀₀	1
0	1	IR _g	R _g	A ₀₁	B ₀₁	R _g =A ₀₁ B ₀₀ /A ₀₀ B ₀₁
1	0	IR _e	R _e	A ₁₀	B ₁₀	R _e =A ₁₀ B ₀₀ /A ₀₀ B ₁₀
1	1	IR _{ge}	R _{ge}	A ₁₁	B ₁₁	R _{ge} =A ₁₁ B ₀₀ /A ₀₀ B ₁₁

遺伝的背景		環境リスク	
I	遺伝的素因は病気発生に対しての影響なし	R _g = 1	環境因子無影響 R _e =1
II	遺伝的素因は病気発生に対しての影響なし	R _g = 1	環境因子による増大 R _e >1
III	遺伝的素因は病気発生を促進する	R _g > 1	環境因子無影響 R _e =1
IV	遺伝的素因は病気発生を促進する	R _g > 1	環境因子による増大 R _e >1
V	遺伝的素因は病気発生を抑制する	R _g < 1	環境因子無影響 R _e =1
VI	遺伝的素因は病気発生を抑制する	R _g < 1	環境因子による増大 R _e >1

I型では遺伝的素因でも環境因子でも片方だけでは病気の発生を促進しませんが、両方だと促進します。例えばフェニルケトン尿症の患者はフェニルアラニン水酸化酵素遺伝子欠損とフェニルアラニン負荷によって発生します。どちらか一方では発生しません。II型では遺伝的素因に関係なく、環境因子により増大します。麻疹などは麻疹ウイルスの感染によって発生しますが、基本的に遺伝的素因は関係ありません。III型では遺伝子のみで発症し、環境は関係ありません。血友病など多くの遺伝病はここに属します。またV、VIのように遺伝的素因が病気の発生を抑制する場合があります。

遺伝子 環境相互作用の疫学研究

病因に関係ある（ありそうな）遺伝子が多型性を持つ場合ではcase control study が強力です。遺伝子多型を病気の危険因子として検討します。遺伝子と環境の相互作用においてOdds ratio (OR_{ge}) は表1のようになります。

例1

Hwang SJ, et al. Association study of transforming growth factor α Taq I polymorphisms and oral clefts: indication of gene-environment interaction in a population-based sample of infants with birth defects. Am J Epi 141:629-36,1995.

著者らは母親の喫煙と TGF α 遺伝子多型と口蓋裂の関連について検討しました。母親に喫煙のみ、あるいは TGF α 遺伝子型のみでは口蓋裂発生との間に有意な相関関係を見出すことができませんでしたが、両者を組み合わせると RR は 5.5 倍、(95% CI: 2.1 – 14.6) と有意に上昇しており、遺伝子 環境因子相互作用の存在が示唆されました。

喫煙	TGF α SNP	口蓋裂	control	risk	OR	95% CI
-	-	36	167	0.215	1.0	
-	+	7	34	0.206	1.0	0.3 – 2.4
+	-	13	69	0.188	0.9	0.4 – 1.8
+	+	13	11	1.182	5.5	2.1 – 14.6
合計		69	281			

問題点

以下の点は一般的臨床研究と基本的には同じですが、遺伝子 環境相互作用という点に特徴的な問題もあるので概説します。

Mis-specification

遺伝子 環境相互作用が存在する時、環境因子のみ、遺伝子だけに調査を限定すると mis-specification につながります。先のお蓋裂の例でも明らかのように、喫煙あるいは TGF α の遺伝子型とお蓋裂との間に有意な相関関係は存在しませんでした。2つの条件が揃ったときにお蓋裂の頻度が5倍以上になりました。

Misclassification of environmental exposures

通常の危険因子を考える以上に exposure の timing に対する配慮が重要です。「成人にとって無害であっても胎児にとっては著しく有害である場合が多い。」などは良い例です。

Misclassification of genotype

個人の DNA より遺伝子型を測定する際、linkage disequilibrium が存在するために misclassification が生じます。Non-differential misclassification であるため、bias は Null に近付きます。また遺伝子型により酵素の活性が異なる場合など、その代謝産物を測定して間接的に遺伝子型を推測する場合がありますが、その関連が必ずしも完全ではなく、やはり bias を生じます。

Confounding

Confounder は遺伝子 環境相互作用を研究する際重要です。特に異なった人種間で比較する場合には、調べようとする遺伝子マーカーや疾患頻度が人種によって異なるため、第三の遺伝子や環境が confounder として働く可能性は十分あります。例えば Pima Indian において Gm3;5;13;14 とインスリン非依存性糖尿病の関連が報告されましたが、実際には白人との混血に多かっただけで、白人との混血によって層化すると Gm3;5;13;14 とインスリン非依存性糖尿病の関連は消失してしまいました。また遺伝子マーカーと疾患の間に相関があるとしても、その遺伝子マーカーとそばの遺伝子が重要であり、その2つの遺伝子の距離が近接しているが故に陰をみているだけの場合もあるのです。

Knowler WC, et al. Gm3,5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. Am J Hum Genet 43:520-26,1988.

用量依存性

N-acetyltransferase 2 (NAT2) の代謝が遅いと膀胱癌の頻度が上がります。また喫煙は NAT2 代謝が遅い場合のみ膀胱癌のリスクファクターとなります。また乳癌でも同様の現象が認められます。Exposure はプラスマイナスのみでなく、量として表してより正確な評価をできる場合があります。

Risch A, et al. Slow N-acetylation genotype is a susceptibility factor in occupational and smoking related bladder cancer risk. Hum Mol Genet 4:231-36,1995.

Ambrosone CB, et al. Cigarette smoking, N-acetyltransferase 2 genetic polymorphisms, and breast cancer risk. JAMA 276:1494-512,1996.

どれくらいのSample 数が必要か？

今まで述べてきたように危険因子への暴露の有無と疾患発生に関与しうる遺伝子変

異によって4つのグループに分類できます。さらにそれぞれのパラメーターは下記のようになります。

危険環境因子への 曝露の有無 (exposure)	疾患発生に関与しうる遺伝子変 異(susceptibility gene)	cases	control	OR
-	-	$[(1-g)(1-e)] / *$	$(1-g)(1-e)$	1.0
-	+	$[g(1-e)R_g] / *$	$g(1-e)$	R_g
+	-	$[e(1-g)R_e] / *$	$e(1-g)$	R_e
+	+	$[geR_{ge}] / *$	ge	R_{ge}

$$a = [(1 - g)(1 - e)] / *$$

$$b = [(1 - g)eR_e] / *$$

$$c = [(1 - e)gR_g] / *$$

$$d = (geR_{ge}) / *$$

e = 対象人口の中でどれくらいの人々が曝露されているか(妊婦の喫煙率など)

g = 対象人口の中でどれくらいの人々が問題視されている遺伝子多型を持つか(TGF α の頻度)

R_e : 遺伝的素因はないが危険環境因子に曝露された人が病気になるリスクを、遺伝子素因もなく危険環境因子にも曝露していない人の病気になるリスクで割ったもの。

R_g : 遺伝的素因を持つが危険環境因子に曝露されていない人が病気になるリスクを、遺伝子素因もなく危険環境因子にも曝露していない人の病気になるリスクで割ったもの。

R_{ge} : 遺伝的素因を持ち危険環境因子にも曝露された人が病気になるリスクを、遺伝子素因もなく危険環境因子にも曝露していない人の病気になるリスクで割ったもの。

$$* = (1 - g)(1 - e) + g(1 - e)R_g + e(1 - g)R_e + geR_{ge}$$

危険因子曝露(exposure)の頻度が極端に高くあるいは低くないかぎり、そして疾患発生を増加させるかもしれない遺伝子変異(susceptible gene)が人々の間で多くみられるときsample数は中等度で十分です。例えば、exposureおよびsusceptible geneが30-70%にみられるようであれば200人の患者数と400人のcontrol数{case/control ratio = 2}があれば80% statistical powerをもって $R_{ge} > 4.0$ を検出することができます。しかしながら例えば乳癌発生に関与するBRCA1 185delAG変異はAshkenazi Jewsの1%に認められ、多くの一般的病気において危険因子の存在は病気発生率をせいぜい2倍に押し上げるに留まることからもわかるとおり、しばしば一般的病気ではexposureおよびsusceptible geneが1-5%です。そのような場合は1,000例以上の検体を集める必要があります。そこでより少ない検体数で遺伝子-環境相互作用を検出できるように以下のような新しい方法が工夫されています。

遺伝子 環境相互作用の新しい臨床研究

I. Case only study

もう一度 TGF α SNP, 口蓋裂を例にとって考えてみましょう。

喫煙	TGF α SNP	口蓋裂	control	risk	OR	95% CI
-	-	36	167	0.215	1.0	
-	+	7	34	0.206	1.0	0.3 – 2.4
+	-	13	69	0.188	0.9	0.4 – 1.8
+	+	13	11	1.182	5.5	2.1 – 14.6
合計		69	281			

$$OR_{CA} = (36 \times 13) / (7 \times 13) = 5.1 \text{ (95\% CI = 1.5 – 18.5)}$$

であり、従来 control と比較した OR = 5.5 と近い値になりました。95% CI の範囲は大分広域になっていることから判る通り、power は落ちますが 281 例の control を必要とせず、69 例のみで遺伝子 環境相互作用の関連を証明できました。しかしながら、case-only studies の考え方は遺伝子 環境因子の間に相乗作用が存在する場合のみ当てはまり、相加的作用の場合には疑問も残ります。

II. Case parental control study

人種間に存在しえる confounder を扱わなくてはならないような場合、case-parental control studies は有効です。また病気と関連する遺伝子多型の頻度が比較的高くその効果も中等度に存在する場合、association study は古典的 linkage analysis より感受性が高い方法です。そこで最近では confounding を考慮に入れ、linkage analysis と association study を融合したような方法が開発されるようになりました。

III. Affected relative-pair study

IV. Twin study

双胎を用いた臨床研究とは、一卵性双胎児(monozygotic twin : MZ) は 100% 同じ遺伝子を共有しているのに対して、二卵性双胎(dizygotic twin: DZ) は一般的に 50% の遺伝子を共有します。MZ と DZ を比較した際、MZ に病気が多ければ遺伝的素因が病気発生に大きく影響すると考え、DZ と MZ の病気の発生率がほとんど同一であれば、環境要素が強いと考えます。

危険環境因子への暴露と遺伝子多型が独立していると仮定して、遺伝子 環境相互作用の存在をスクリーニングするために開発された方法です。もし遺伝子 環境相互作用の存在を知るだけであれば、control を用いる必要はありません。基本的には以下の 2x2 table で表されます。

	遺伝的素因なし	遺伝的素因あり
暴露あり	a	b
暴露なし	c	d

$$a = [(1 - g)(1 - e)] / *$$

$$b = [(1 - g)eRe] / *$$

$$c = [(1 - e)gRg] / *$$

$$d = (geR_{ge}) / *$$

e = 対象人口の中でどれくらいの人々が暴露されているか（妊婦の喫煙率など）

g = 対象人口の中でどれくらいの人々が問題視されている遺伝子多型を持つ（TGF α の頻度）

R_e: 遺伝的素因はないが危険環境因子に暴露された人が病気になるリスクを、遺伝子素因もなく危険環境因子にも暴露していない人の病気になるリスクで割ったもの。

R_g: 遺伝的素因を持つが危険環境因子に暴露されていない人が病気になるリスクを、遺伝子素因もなく危険環境因子にも暴露していない人の病気になるリスクで割ったもの。

R_{ge}: 遺伝的素因を持ち危険環境因子にも暴露された人が病気になるリスクを、遺伝子素因もなく危険環境因子にも暴露していない人の病気になるリスクで割ったもの。

$$* = (1 - g)(1 - e) + g(1 - e)R_g + e(1 - g)R_e + geR_{ge}$$

$$OR_{CA} = R_{ge} / (R_e \times R_g) \times OR_{CO}$$

OR_{CA} = case only odds ratio

OR_{CO} = 暴露と遺伝素因に関係するcontrol におけるodds ratio