

# Analysis of Complex Traits

1. Monogenic disease と Complex disease の相違を述べてください。

**Monogenic disease** とはメンデルの法則に則って遺伝する疾患で、通常は1つの遺伝子の変異により引き起こされます。Complex disease と比べ希なことが多く、genetic homogeneity のため比較的早期に発症する傾向にあります。そして、penetrance も高く、segregation analysis を用いた家系解析により多くの病気の原因遺伝子異常が発見されてきました。またparametric linkage analysis もしばしば用いられます。

**Complex disease** とはメンデルの法則では説明できない遺伝する疾患で、病気が病気でないかといったyes/no よりも、血圧のように連続した数値として表すことができます。複数の遺伝子が関与し、遺伝子同志、遺伝子 - 環境因子相互作用によって病気の性質が左右されるため、その病因は複雑です。そのため遺伝子異常と症状の間にギャップがみられることが多く、遺伝子異常をもつからといって症状を呈するとは限らず (incomplete genetic penetrance)、逆に遺伝子異常を示さないのに同じ病気を発症することがあります (phenocopy)。ミトコンドリア病、imprinting, anticipation など特殊な遺伝形式も complex disease に含まれます。Sib-pair studies や affected pedigree member studies などの non-parametric linkage analytic methods が病因遺伝子を同定するために用いられます。しかし、family-based association studies の方がより強力です。遺伝子 遺伝子相互作用や遺伝子 - 環境相互作用に関しては case-control study などの疫学的手法を用います。

遺伝疫学の定義とは？

遺伝疫学 (genetic epidemiology) とは、病気の遺伝因子について考察する学問であり、遺伝子 - 遺伝子、遺伝子 - 環境相互作用を検討します。古典的疫学に対し遺伝疫学は遺伝因子を考慮し、家族など遺伝子を共有する人は似ていると考える点が特徴です。Population genetics とは病気に焦点をあてるか否かの相違があります。一方純粋な遺伝学に対して遺伝疫学は病因、分布、population レベルでの疾病コントロールに重点を置きます。

Complex disease を吟味するのに用いる遺伝疫学的手法について解説してください。

## (1) Complex disease gene mapping

Transmission disequilibrium test (TDT) などの遺伝疫学的手法が linkage disequilibrium の存在を証明するには有効です。Complex disease でしばしばそうなのですが、疾患遺伝子による疾患リスクがそれほど高くない場合には TDT は古典的 linkage analysis より強力です。

## (2) Gene-Disease Association

Case-control study がしばしば用いられます。この場合、ある遺伝子 polymorphism を持っていることが exposure にあたります。最近では、民族による遺伝子 polymorphism の

偏りと疾患の偏りをモス日つけようとする研究も目立ちます。

(3) Gene-Environment interaction

Cohort, case-control, cross-sectional studies などの遺伝疫学的手法が complex disease の遺伝 - 環境相互作用を検討する際しばしば用いられます。

(4) Pharmacogenetics

いくつかの遺伝子が環境因子と相互作用を起こして、薬剤に対する反応性を規定します。薬剤の反応性を規定する因子は非常に多くあるため、反応性は 0 or 1 でなく連続的なものとなります。Family based association study などが薬剤反応性に関与する個々の variation を調べるのに適しています。

2. あなたは complex disease の中から高血圧を選び、1,000 の核家族（父母、子供 2 人の構成）について高血圧の有無を調査しました。この対象を含む一般人口の高血圧の割合は 10% と仮定してください。

兄弟 1	兄弟 2	家族数
0	0	810
0	1	90
1	0	70
1	1	30

0 = 高血圧なし、1 = 高血圧あり

この集団における高血圧の  $\lambda_s$  はいくつですか？この集団において高血圧は monogenic ですか、それとも complex disease ですか？この高血圧に關与している遺伝子を研究するために、あなたは次のステップとしてどのような方法をとりますか？

あなたは 20 の核家族における収縮期血圧の familial segregation study を行うことにしました。これらの家族で収縮期血圧の familial aggregation を認めますか？次にどのような調査を行いますか？

収縮期血圧 (mmHg)

家族 ID	父親	母親	第 1 子	第 2 子
1	120	100	110	112
2	85	104	100	103
3	90	100	94	96
4	113	113	99	111
5	130	105	115	120
6	109	116	110	112
7	90	110	100	103
8	100	93	94	96
9	113	96	99	111
10	125	105	115	120
11	105	113	110	112
12	100	105	101	103
13	90	97	94	96
14	113	96	99	111
15	122	123	115	120
16	119	109	110	112
17	95	105	99	101
18	90	100	93	94
19	113	96	98	108
20	130	105	116	121

第1子が高血圧で、第2子も高血圧である条件付き確率は

$$30 / (70 + 30) = 0.3$$

一般の高血圧の頻度は 10%、  
よって

$$\lambda s = 0.3/0.1 = 3.0$$

逆に第2子を軸に考えた場合、 $30 / (90 + 30) = 0.25$

もしも高血圧が常染色体優性遺伝であれば 50%に発現していなくてはならず、常染色体劣性であれば 25%に発現していなくてはおかしいこととなります。しかも、一般の高血圧浸透率が 10%ということは、兄弟で病気を共有する確率が 25 - 30 とはいっても、monogenic disease を思わせる程高い数値ではありません。よって、この population における高血圧の要因として complex disease が考えられます。

次は環境と遺伝要因がどれくらいの比率で関与しているかを調べるべきでしょう。なぜなら  $\lambda s = 3.0$  or  $2.5$  は遺伝要素を示唆しますが、遺伝病であることを意味するわけではないのです。まずは twin study (or adaptation study)ができれば、この方法が環境・遺伝要因の比率を知る上で最も意義ある情報を与えてくれるでしょう。そして、「遺伝性要素がその後の調査をすすめるに足るほどありそうだ」ということになれば、(1) できるだけ多くの高血圧家系を集め、兄弟例とその両親から検体を取り、genome wide screening を行う、(2)逆に高血圧症例と高血圧を持たない兄弟の組みあわせ(discordant sib pairs)を集め、genome screening を行う、(3) 高血圧に関して family based TDT study を行う、などが考えられます。さらに、高血圧の中でも腎性高血圧あるいは本態性高血圧など、ある一定の臨床的特徴をもつ亜集団(intermediate phenotype)を選び出し、monogenic disease に行ったように linkage and association studies を行うのも一方です。

収縮期血圧と家族内集積について検討します。まずはそれぞれの組み合わせで相関係数と P 値を求めます。例えば父と母の間に高血圧の相関が弱く、親と子の間で強く、兄弟間は中間であるとすれば遺伝性を強く疑えるからです。

```
. correlate father mother first second  
(obs=20)
```

	father	mother	first	second
father	1.0000			
mother	0.2254	1.0000		
first	0.7919	0.5841	1.0000	
second	0.9081	0.4752	0.9010	1.0000

```
. regress father mother
```

Source	SS	df	MS	Number of obs =	20
Model	200.458032	1	200.458032	F( 1, 18) =	0.96
Residual	3746.34197	18	208.130109	Prob > F =	0.3394
				R-squared =	0.0508

```
-----+-----
Total | 3946.80 19 207.726316
Adj R-squared = -0.0019
Root MSE = 14.427
```

```
-----+-----
father | Coef. Std. Err. t P>|t| [95% Conf. Interval]
-----+-----
mother | .4198953 .4278552 0.981 0.339 -.478995 1.318786
_cons | 63.69994 44.84843 1.420 0.173 -30.5231 157.923
-----+-----
```

```
. regress father first
```

```
Source | SS df MS
-----+-----
Model | 2475.14291 1 2475.14291
Residual | 1471.65709 18 81.7587275
Total | 3946.80 19 207.726316
Number of obs = 20
F( 1, 18) = 30.27
Prob > F = 0.0000
R-squared = 0.6271
Adj R-squared = 0.6064
Root MSE = 9.0421
```

```
-----+-----
father | Coef. Std. Err. t P>|t| [95% Conf. Interval]
-----+-----
first | 1.410016 .2562661 5.502 0.000 .8716214 1.948411
_cons | -38.40721 26.61327 -1.443 0.166 -94.3196 17.50519
-----+-----
```

```
. regress father second
```

```
Source | SS df MS
-----+-----
Model | 3254.85305 1 3254.85305
Residual | 691.946946 18 38.441497
Total | 3946.80 19 207.726316
Number of obs = 20
F( 1, 18) = 84.67
Prob > F = 0.0000
R-squared = 0.8247
Adj R-squared = 0.8149
Root MSE = 6.2001
```

```
-----+-----
father | Coef. Std. Err. t P>|t| [95% Conf. Interval]
-----+-----
second | 1.501454 .1631724 9.202 0.000 1.158642 1.844267
_cons | -54.70723 17.69333 -3.092 0.006 -91.87954 -17.53492
-----+-----
```

```
. regress mother first
```

```
Source | SS df MS
-----+-----
Model | 387.931646 1 387.931646
Residual | 749.018354 18 41.6121308
Total | 1136.95 19 59.8394737
Number of obs = 20
F( 1, 18) = 9.32
Prob > F = 0.0068
R-squared = 0.3412
Adj R-squared = 0.3046
Root MSE = 6.4507
```

```
-----+-----
mother | Coef. Std. Err. t P>|t| [95% Conf. Interval]
-----+-----
first | .5582152 .1828243 3.053 0.007 .1741155 .9423149
_cons | 46.74682 18.98633 2.462 0.024 6.858013 86.63562
-----+-----
```

```
. regress mother second
```

```
Source | SS df MS
-----+-----
Model | 256.794023 1 256.794023
Residual | 880.155977 18 48.8975543
Total | 1136.95 19 59.8394737
Number of obs = 20
F( 1, 18) = 5.25
Prob > F = 0.0342
R-squared = 0.2259
Adj R-squared = 0.1829
Root MSE = 6.9927
```

```
-----+-----
mother | Coef. Std. Err. t P>|t| [95% Conf. Interval]
-----+-----
second | .4217343 .1840306 2.292 0.034 .0351004 .8083683
_cons | 58.96052 19.95506 2.955 0.008 17.03649 100.8846
-----+-----
```

```

-----
. regress first second

Source |         SS          df           MS                Number of obs =      20
-----+-----
Model |    1010.5433          1    1010.5433                F( 1, 18) =    77.60
Residual |    234.406705        18    13.0225947                Prob > F      =    0.0000
-----+-----
Total |    1244.95          19    65.5236842                R-squared     =    0.8117
                                           Adj R-squared =    0.8013
                                           Root MSE    =    3.6087

-----
first |         Coef.      Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
second |    .8366117      .0949719     8.809  0.000     .6370831     1.03614
_cons  |   13.11227     10.29813     1.273  0.219    -8.52329     34.74784
-----

```

上の解析結果から明らかなように、父 - 母の組み合わせは有意ではありませんが、親が高血圧を持っているとその子供も高血圧を持つ傾向にありました。つまり遺伝的因子を疑うに十分な結果といえます。更に、年齢、性別、肥満度などの高血圧と関連する因子も含めて収縮期血圧を adjustment するとより正確なデータを得ることができます。

上の結果解釈をさらに確かなものとするには twin study が適しています。1 卵性双胎での高血圧併発率の方が、2 卵性双胎高血圧併発率あるいは兄弟間高血圧併発率よりも高ければ遺伝性をさらに強く疑うことができます。一方、この 2 群が同じであれば環境因子を疑うべきでしょう。得てして兄弟は環境を共有するからです。さらに、何かの社会的事情で血縁ではない子供をある家庭で引き取ったとします(adaptation)。そのような子供と、一緒に暮らした兄弟、あるいは 1 卵性双胎間での高血圧併発率が等しければ、より強く環境因子を病因として考えることができます。

Segregation analysis も major locus がるかどうか、あるいはメンデル遺伝形式に当てはまるかどうかを調べるには適しています。

Genome screening study は兄弟共になが高血圧を持つ組(high concordant sibpairs)、兄弟共に高血圧を持たない組(low concordant sibpairs)、兄弟で高血圧をもつものともたないものが混在する組(extremely discordant sibpairs)、を用いて行われます。

Family-based association studies (TDT, SDT, sib-TDT)、あるいは Case-control study が行われてもいいかもしれません。特に遺伝子多型と環境因子の関係は疫学的調査のよい対象です。

3. 何故 segregation analysis が monogenic disease を行なう上で適当であり、complex traits を検討する際、必ずしも必要ないのですか？

Segregation analysis は元来観察された家族の遺伝データが単純なメンデルの法則に従うかどうかについてデザインされた方法です。最近では、遺伝解析がより詳細に行えるようになったため complex trait の intermediate trait に対しても行われるようになりましたが、基本的にはメンデル法則に当てはまるかどうかを検討するものであることは変わりありません。

遺伝性に関する評価法の1つに Falconer's estimate があり下記の公式で示されます。

$$H^2 = 2(r_{MZ} - r_{DZ})$$

この公式を使って下記のデータを解析し、結果について考察を加えてください。

Twin id	1 卵性双胎 = 1 2 卵性双胎 = 0	収縮機血圧 mmHg
1	1	120
1	1	116
2	1	101
2	1	97
3	0	106
3	0	118
4	1	135
4	1	129
5	1	107
5	1	105
6	1	120
6	1	124
7	0	100
7	0	112
8	0	132
8	0	120
9	0	136
9	0	142
10	0	134
10	0	110

ANOVA F test と同様に双胎内の variance に比べ双胎間の variance が大きい場合有意と考えます。

```
. sort twin
. by twin: sum sbp

-> twin=      0
Variable |      Obs      Mean  Std. Dev.      Min      Max
-----+-----
sbp |      10      121  14.27508      100      142

-> twin=      1
Variable |      Obs      Mean  Std. Dev.      Min      Max
-----+-----
sbp |      10     115.4  12.51843      97      135

. sort id
. by id: sum sbp
-> id=      1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
sbp	2	118	2.828427	116	120
-> id= 2					
Variable	Obs	Mean	Std. Dev.	Min	Max
sbp	2	99	2.828427	97	101
-> id= 3					
Variable	Obs	Mean	Std. Dev.	Min	Max
sbp	2	112	8.485281	106	118
-> id= 4					
Variable	Obs	Mean	Std. Dev.	Min	Max
sbp	2	132	4.242641	129	135
-> id= 5					
Variable	Obs	Mean	Std. Dev.	Min	Max
sbp	2	106	1.414214	105	107
-> id= 6					
Variable	Obs	Mean	Std. Dev.	Min	Max
sbp	2	122	2.828427	120	124
-> id= 7					
Variable	Obs	Mean	Std. Dev.	Min	Max
sbp	2	106	8.485281	100	112
-> id= 8					
Variable	Obs	Mean	Std. Dev.	Min	Max
sbp	2	126	8.485281	120	132
-> id= 9					
Variable	Obs	Mean	Std. Dev.	Min	Max
sbp	2	139	4.242641	136	142
-> id= 10					
Variable	Obs	Mean	Std. Dev.	Min	Max
sbp	2	122	16.97056	110	134

MSAMZ = sum of squares among monozygous twin,  
 $MZ/(nMZ-1) = 1366.4/4 = 341.6$

MSWMZ = sum of squares within monozygous twin,  
 $MZ/nMZ = 44/5 = 8.8$

MSADZ = sum of squares among dizygous twin,  
 $DZ/(nDZ-1) = 1312/4 = 328.0$

MSWDZ = sum of squares within dizygous twin,  
 $DZ/nDZ = 522/5 = 104.4$



$$rMZ = (MSAMZ - MSWMZ) / (MSAMZ + MSWMZ) = (341.6 - 8.8) / (341.6 + 8.8) = 0.95$$

$$rDZ = (MSADZ - MSWDZ) / (MSADZ + MSWDZ) = (328.0 - 104.4) / (328.0 + 104.4) = 0.517$$

$$h^2 = 2(rMZ - rDZ) = 2(0.95 - 0.517) = 0.866$$

よって血圧の程度の差に占める遺伝的要素は 86.6% であると結論できます。

4. Intermediate phenotype について定義し、何故 complex disease gene mapping を考える際、この intermediate phenotype を考えることが重要なのか論じてください。

基本的に complex disease では同じ表現型でもいくつかの病因が存在することがしばしばです。例えば高血圧という表現型は同じですが、腎性のものであれば内分泌性のもでもあります。しかし、臨床的に検討すると病因によって症状や検査所見が異なります。例えば高血圧で腎動脈の狭窄とこれによるホルモン動態を示すものを腎性高血圧としますと、これは intermediate phenotype にあたります。このように疾患亜分類に焦点をあてることにより遺伝性をより鮮明に検討することができます。遺伝性の程度は多々ある高血圧の病因で当然異なるでしょう。

5. Angiotensin の polymorphism; M235T を、1000 人を対象に調査したところ、MM は 54 人、MT は 300 人、TT は 646 人でした。Allele frequencies と期待される 3 つの genotypes を計算してください。また  $\chi^2$  を用いて genotype distribution が Hardy-Weinberg equilibrium (HWE) に合っているか検討してください。自由度はいくつですか？

まずは M235T の意味は何ですか？  
アミノ酸記号

Amino acid		
alanine	Ala	A
arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

アンジオテンシンの 235 番目のアミノ酸メチオニンがスレオニンに置換した変異を意

味します。いわゆる single nucleotide polymorphism (SNP)です。染色体は2本対になっているわけで、アンギオテンシンの2つの染色体でつくられています。両方ともメチオニンであった人が54人、メチオニンとスレオニンの混合が300人、両方ともスレオニンであった人が646人ということになります。

M allele frequency;  $p = (54 + 300/2)/1000 = 0.204$

T allele frequency;  $q = 1 - p = 0.796$

よって期待される3つの遺伝子型の数は

MM:  $1000p^2 = 41.6$

MT:  $10002pq = 324.8$

TT:  $1000q^2 = 633.6$

$(54 - 41.6)^2/41.6 + (300 - 324.8)^2/324.8 + (646 - 633.6)^2/633.6 = 5.8324$

$df = 3 - 1 - 1 = 1$

$p < 0.05$

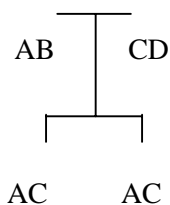
Hardy Weinberg equilibrium の法則には合わないと結論できます。

あなたは収縮機血圧(sbp) に関するsib pair linkage study の解析を依頼されました。そこで、あなたはAngiotensin I converting enzyme (ACEI) に関してmicrosatellite marker を使って検討しました。このマーカーは4種類あり、それぞれの一般人口の頻度は $A_1 = 0.20$ ,  $A_2 = 0.25$ ,  $A_3 = 0.10$ ,  $A_4 = 0.05$ ,  $A_5 = 0.27$ ,  $A_6 = 0.13$  です。そして、調査した12組の兄弟例のデータは以下のようです。

家族 ID	性別 男 = 1、女 = 2	マーカー	マーカー	収縮期血圧 (sbp)
1	1	1	2	158
1	2	1	2	158
2	2	1	2	149
2	1	1	5	154
3	1	2	4	150
3	2	3	6	165
4	1	1	3	172
4	1	1	6	176
5	2	2	2	156
5	1	2	5	162
6	2	5	5	157
6	1	5	5	158
7	1	1	1	151
7	2	3	3	162
8	1	1	3	155
8	1	3	4	156
9	2	2	2	165
9	2	2	2	166
10	2	1	5	160
10	2	5	6	166
11	2	4	5	158
11	2	5	6	165
12	1	5	5	162
12	2	1	5	155

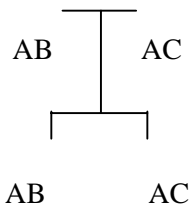
それぞれの兄弟における IBD を共有する allele の頻度を検討してください。収縮期血圧とこのマーカーの間にリンクがあるかどうか Haseman-Elston method を用いて検討してください。

そもそも identical by descent (IBD)とは何でしょう？  
 血縁内で全く同じ遺伝子のパターンを受け継いでいることを指します。  
 下の場合何個同じパターンを引き継いでいますか？



IBD = 2

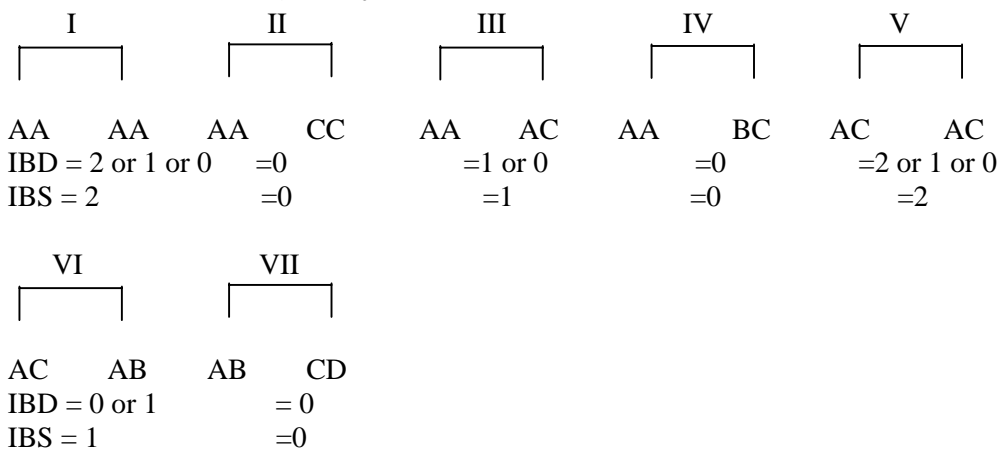
2人の子供のAは父親、Cは母親からきています。



IBD = 0

子供の子供のBは父親、Cは母親なので男の子のAは母親、女の子のAは父親由来です。この場合のマーカーAは同じでもこの遺伝子自体は異なると考えます。よってIBDは0です。一方 identical by state (IBS)は1です。IBSはマーカーの一致数でみます。IBD=<IBSです。

もしも親の遺伝子マーカー解析資料が無ければどうでしょうか？IBDを言い当てることはできません。そこで、そのマーカーの一般人口における比率を参考に算出します。その公式は以下のようなものです。



	IBD が 0 である確率	IBD が 1 である確率	IBD が 2 である確率
I	$P_A^4$	$P_A^3$	$P_A^2$
II	$2P_A^2P_C^2$	0	0
III	$4P_A^3P_C$	$2P_A^2P_C$	0
IV	$4P_A^3P_CP_B$	0	0
V	$4P_A^2P_C^2$	$P_AP_C(P_A + P_C)$	$2P_AP_C$
VI	$8P_A^3P_BP_C$	$2P_AP_BP_C$	0
VII	$P_AP_BP_CP_D$	0	0

$$\pi = [f_2 * (\text{IBD が 2 である確率}) + f_1 * (\text{IBD が 1 である確率} / 2)] / [f_2 * (\text{IBD が 2 である確率}) + f_1 * (\text{IBD が 1 である確率} / 2) + f_0 * (\text{IBD が 0 である確率})]$$

fi: i 個の IBD を親等が共有する確率;  $f_2 = 1/4$ ,  $f_1 = 1/2$ ,  $f_0 = 1/4$

家族 ID	Y ( 血圧の差 ) <sup>2</sup>	$\pi$
(1)	0	0.79
(2)	25	0.36
(3)	225	0.00
(4)	16	0.36
(5)	36	0.40
(6)	1	0.79
(7)	121	0.00
(8)	1	0.42
(9)	1	0.80
(10)	36	0.32
(11)	49	0.32
(12)	49	0.39

$$(y_{1j} - y_{2j})^2 = \alpha + \beta\pi$$

$H_0: \beta=0$  versus  $H_A: \beta < 0$

```
. regress Y X
```

Source	SS	df	MS	Number of obs = 12		
Model	29503.8212	1	29503.8212	F( 1, 10) =	16.48	
Residual	17906.8454	10	1790.68454	Prob > F =	0.0023	
Total	47410.6667	11	4310.06061	R-squared =	0.6223	
				Adj R-squared =	0.5845	
				Root MSE =	42.316	

Y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
X	-192.1823	47.34605	-4.059	0.002	-297.6758	-86.6887
_cons	125.9419	23.03593	5.467	0.000	74.6146	177.2691

$$E(Y) = 125.5 - 192.2 \pi, p=0.002$$

よって血圧はこの Angiotensin converting enzyme のマーカーと有意にリンクすると結論できません。