

- 目的: マルチアレル多型におけるHWE平衡からの逸脱を評価する。
 適用例1: 単一ローカスに複数の多型が存在する場合。HLA遺伝子など
 適用例2: CNV多型 (別途染色体毎のコピー数推定が必要)

○方法

- ・1 locus における kアレル多型で構成された、Nサンプルに対するディプロタイプデータを対象とする。
 - ・アレル i の観測頻度を、 $f_{(i)}$ とする。(i=1...k)
 - ・アレル i, j で構成されるディプロタイプの観測度数を O_{ij} 、期待度数を E_{ij} とする。
- 下記定義式に従い、HWE χ^2 統計量(= X_{HWE}) 及び対応するP値(= P_{HWE}) を求める。

$$E_{ij} = 2Nf_{(i)}f_{(j)} \quad (i \neq j),$$

$$Nf_{(i)}f_{(j)} \quad (i = j).$$

$$X_{HWE} = \sum_{i=1}^k \sum_{j=i}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}.$$

$$X_{HWE} \sim \chi^2(df), \quad df = \left(\frac{k(k+1)}{2} - 1 \right) - (k-1) = \frac{k(k-1)}{2}.$$

- ・ハプロタイプレベルでのパーミュテーションを行い、各ステップでの X_{HWE} を得る。
- ・得られた分布を、帰無仮説下 (HWE平衡からの逸脱が無い状態) での X_{HWE} の分布と仮定し、パーミュテーションP値(= $P_{HWE, Perm}$)を求めらる。
- ・得られた分布と、 $\chi^2_{(df)}$ との合致を、適合度検定で評価する。

○結果

①: 3アレル多型データに対して

ディプロタイプ	O_{ij}	E_{ij}	$(O-E)^2/E$
A/A	210	202.01	0.316
A/B	580	619.08	2.467
A/C	1010	986.91	0.540
B/B	500	474.32	1.390
B/C	1500	1512.28	0.100
C/C	1200	1205.41	0.024
sum	5000	5000	4.838

アレル	$f_{(i)}$
A	0.201
B	0.308
C	0.491
sum	1

k=3, df=3

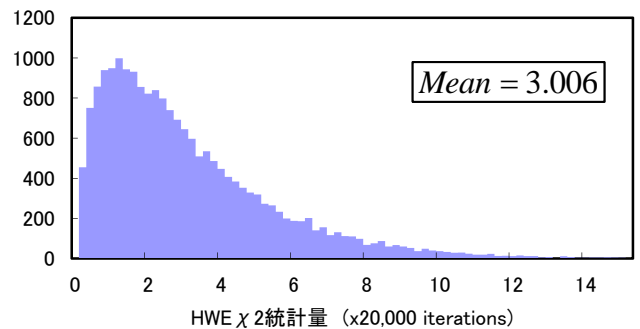
$X_{HWE}=4.838, P_{HWE}=0.1840$

$P_{HWE, Perm}=0.1829$ (x100,000 iterations)

適合度検定P値=1

⇒パーミュテーションの結果、df=3の χ^2 分布に合致した分布が得られた。
 HWE χ^2 統計量から得たP値と、パーミュテーションで得たP値がほぼ同一の値となった。

図1 : 3 alleles



②: 2410名のRA検体におけるHLA-DRB1アレルに対して

アレルNo.	観測数	$f_{(i)}$	アレルNo.	観測数	$f_{(i)}$
1	1208	25.1%	19	53	1.1%
2	827	17.2%	20	52	1.1%
3	414	8.6%	21	35	0.7%
4	321	6.7%	22	12	0.2%
5	277	5.7%	23	12	0.2%
6	272	5.6%	24	10	0.2%
7	193	4.0%	25	10	0.2%
8	163	3.4%	26	5	0.1%
9	123	2.6%	27	3	0.1%
10	123	2.6%	28	2	0.0%
11	117	2.4%	29	2	0.0%
12	116	2.4%	30	1	0.0%
13	109	2.3%	31	1	0.0%
14	101	2.1%	32	1	0.0%
15	74	1.5%	33	1	0.0%
16	68	1.4%	34	1	0.0%
17	58	1.2%	35	1	0.0%
18	54	1.1%	sum	4820	100%

k=35, df=595

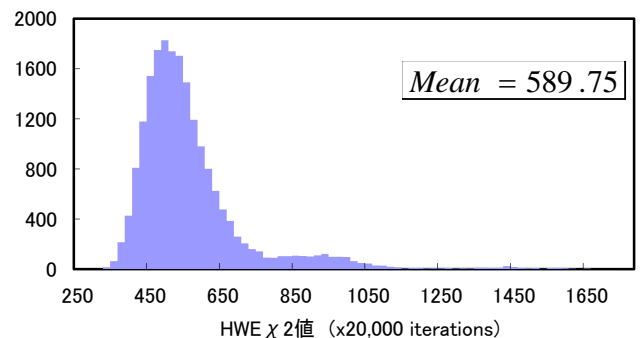
$X_{HWE}=661.68, P_{HWE}=0.0298$

$P_{HWE, Perm}=0.1471$ (x100,000 iterations)

適合度検定P値=0

⇒パーミュテーションの結果、df=595の χ^2 分布と合致しない分布が得られた。
 HWE χ^2 統計量から得たP値と、パーミュテーションで得たP値に差を認めた。

図2 : 35DRB1 alleles in 2410 RA cases



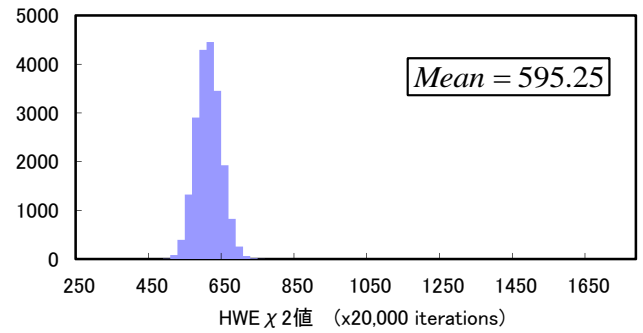
* : ②例では、期待度数の低いディプロタイプが多く存在し、漸近近似が不適切であった可能性が考えられた。

⇒同じアレル数で、アレル頻度が均等なデータにおける、 X_{HWE} の分布を評価した。

k=35, df=595
適合度検定P値=1

⇒df=595の χ^2 分布に合致した分布が得られた。
⇒図2における χ^2 分布からの逸脱の要因として
低ディプロタイプ期待頻度の影響が考えられた。

図3 : 35 alleles with same freq. in 2450 samples



○考察

- ・ディプロタイプ期待頻度が低い場合、通常のHWE χ^2 値は想定される自由度の χ^2 分布からの乖離を生じると考えられた。
- ・パーミュテーションを用いることにより、期待度数が低い場合においてHWE P値をより適切に求めることができると考えられた。

○分布偏移の要因

・図2における分布の偏移(I~III)と、低期待頻度との関連

- I : 平均値の偏移・・・実質的な自由度が引低くなるため
- II : 第2/3ピークの出現・・・期待度数の低いセルに高い観測度数が含まれる場合、 χ^2 値がinflationするため
- III : 分散の増大・・・I+II(？)

図2 : 35DRB1 alleles in 2410 RA cases

