

Whole genome imputation実施後ジェノタイプデータに対する関連解析方針案 (ver1.1)

2011/12/10 理化学研究所 統計解析研究チーム 岡田随象

①: 概要

- Whole genome imputation実施後ジェノタイプデータに対するケースコントロール解析実施手順案を記載する。
- ImputationにはMACH (<http://www.sph.umich.edu/csg/abecasis/MACH/index.html>) を使用した場合を想定している。

②: Imputation実施後ジェノタイプデータに対する関連解析の特徴

- 通常のGWASデータ解析では、各SNPのジェノタイプは A/A, A/B, B/B の3ジェノタイプで表記される。
- additive modelを仮定し、A/A⇒0, A/B⇒1, B/B⇒2 と、3整数で表現されるアレル本数に変換して解析に用いることが多い。
- Imputation実施後ジェノタイプデータでは、各ジェノタイプ別の事後確率もしくは0~2の範囲の小数で表現されるアレル本数の期待値 (allele dosage) で与えられる点異なる。
- 本資料では、allele dosageに基づく解析方針案を記載する。
- 解析ソフトには、MACHの出力ファイルに対応した"mach2dat"を採用する。(MACHと同じhomepage上で提供されている。)

③: 【手順1】 MACH出力ファイルの変換

- MACH出力ファイルのうち、hoge.mlinfo 及び hoge.mldose を mach2dat の入力ファイルとして使用する。
- hoge.mlinfoには各SNPの情報(rsID, imputationスコア, アレル頻度、etc.)が記載されている。
- hoge.mldose には各サンプル、各SNP別の allele dosage が記載されている。
- hoge.mldose 上の allele dosage は、hoge.mlinfo 上のアレル"A1"のアレル本数を表している。
- Imputation実施時に入力ファイル上で指定したアレル1/2と、アレルA1/AI2の対応はSNP毎に異なる。
- hoge.mlinfo, hoge.mldose をそのまま mach2dat の入力ファイルに使用することも可能であるが、予めアレル1/2とアレルA1/AI2を対応させておけば、hoge.mlinfo を逐次参照する必要がなくなる。そのため、以下のような変換を行う。

~hoge.mlinfo, hoge.mldose → hoge_flip.mlinfo, hoge_flip.mldoseへの変換~
(網掛け部分が変換対象)

| hoge.mlinfoファイル | | | | | | | hoge_flip.mlinfoファイル | | | | | | |
|-----------------|----|----|--------|--------|---------|--------|----------------------|----|----|--------|--------|---------|--------|
| SNP | A1 | A2 | Freq1 | MAF | Quality | Rsq | SNP | A1 | A2 | Freq1 | MAF | Quality | Rsq |
| rs000001 | 1 | 2 | 0.9853 | 0.0147 | 0.9708 | 0.0126 | rs000001 | 2 | 1 | 0.0147 | 0.0147 | 0.9708 | 0.0126 |
| rs000002 | 2 | 1 | 0.9956 | 0.0044 | 0.9912 | 0.0752 | rs000002 | 2 | 1 | 0.9956 | 0.0044 | 0.9912 | 0.0752 |
| rs000003 | 2 | 1 | 0.9792 | 0.0208 | 0.9589 | 0.0556 | rs000003 | 2 | 1 | 0.9792 | 0.0208 | 0.9589 | 0.0556 |
| rs000004 | 2 | 1 | 0.9937 | 0.0063 | 0.9875 | 0.0033 | rs000004 | 2 | 1 | 0.9937 | 0.0063 | 0.9875 | 0.0033 |
| rs000005 | 1 | 2 | 0.9943 | 0.0057 | 0.9886 | 0.005 | rs000005 | 2 | 1 | 0.0057 | 0.0057 | 0.9886 | 0.005 |
| rs000006 | 1 | 2 | 0.9263 | 0.0737 | 0.8581 | 0.0285 | rs000006 | 2 | 1 | 0.0737 | 0.0737 | 0.8581 | 0.0285 |
| rs000007 | 2 | 1 | 0.907 | 0.093 | 0.8227 | 0.0325 | rs000007 | 2 | 1 | 0.907 | 0.093 | 0.8227 | 0.0325 |
| rs000008 | 2 | 1 | 0.907 | 0.093 | 0.8226 | 0.0332 | rs000008 | 2 | 1 | 0.907 | 0.093 | 0.8226 | 0.0332 |
| rs000009 | 1 | 2 | 0.7305 | 0.2695 | 0.5526 | 0.0697 | rs000009 | 2 | 1 | 0.2695 | 0.2695 | 0.5526 | 0.0697 |

hoge.mldoseファイル

| Sample | ML_DOSE | 1.973 | 1.994 | 1.984 | 1.986 | 1.986 | 1.846 | 0.798 | 1.797 | 1.397 |
|----------------------|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Sample001->Sample001 | ML_DOSE | 1.973 | 1.994 | 1.984 | 1.986 | 1.986 | 1.846 | 0.798 | 1.797 | 1.397 |
| Sample002->Sample002 | ML_DOSE | 1.966 | 1.995 | 1.005 | 1.992 | 1.993 | 1.739 | 1.889 | 1.89 | 1.606 |
| Sample003->Sample003 | ML_DOSE | 1.976 | 1.994 | 0.924 | 1.986 | 1.98 | 1.773 | 1.838 | 1.838 | 1.509 |
| Sample004->Sample004 | ML_DOSE | 1.983 | 1.996 | 1.947 | 1.993 | 0.994 | 1.805 | 1.904 | 1.905 | 1.613 |
| Sample005->Sample005 | ML_DOSE | 1.974 | 0.994 | 1.985 | 1.981 | 1.988 | 1.882 | 1.758 | 1.757 | 1.391 |
| Sample006->Sample006 | ML_DOSE | 1.926 | 1.994 | 1.986 | 1.989 | 1.99 | 1.913 | 1.795 | 1.794 | 1.434 |
| Sample007->Sample007 | ML_DOSE | 1.974 | 1.994 | 1.986 | 1.97 | 1.987 | 1.871 | 1.811 | 1.811 | 1.435 |
| Sample008->Sample008 | ML_DOSE | 1.973 | 1.994 | 0.125 | 1.972 | 1.987 | 0.897 | 1.804 | 1.804 | 1.397 |
| Sample009->Sample009 | ML_DOSE | 1.983 | 0.995 | 1.963 | 1.984 | 1.992 | 1.814 | 0.888 | 1.888 | 1.65 |
| Sample010->Sample010 | ML_DOSE | 1.974 | 1.994 | 1.984 | 1.981 | 1.987 | 1.871 | 1.789 | 1.789 | 1.396 |

↓

hoge_flip.mldoseファイル

| Sample | ML_DOSE | 0.027 | 1.994 | 1.984 | 1.986 | 0.014 | 0.154 | 0.798 | 1.797 | 0.603 |
|----------------------|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Sample001->Sample001 | ML_DOSE | 0.027 | 1.994 | 1.984 | 1.986 | 0.014 | 0.154 | 0.798 | 1.797 | 0.603 |
| Sample002->Sample002 | ML_DOSE | 0.034 | 1.995 | 1.005 | 1.992 | 0.007 | 0.261 | 1.889 | 1.89 | 0.394 |
| Sample003->Sample003 | ML_DOSE | 0.024 | 1.994 | 0.924 | 1.986 | 0.02 | 0.227 | 1.838 | 1.838 | 0.491 |
| Sample004->Sample004 | ML_DOSE | 0.017 | 1.996 | 1.947 | 1.993 | 1.006 | 0.195 | 1.904 | 1.905 | 0.387 |
| Sample005->Sample005 | ML_DOSE | 0.026 | 0.994 | 1.985 | 1.981 | 0.012 | 0.118 | 1.758 | 1.757 | 0.609 |
| Sample006->Sample006 | ML_DOSE | 0.074 | 1.994 | 1.986 | 1.989 | 0.01 | 0.087 | 1.795 | 1.794 | 0.566 |
| Sample007->Sample007 | ML_DOSE | 0.026 | 1.994 | 1.986 | 1.97 | 0.013 | 0.129 | 1.811 | 1.811 | 0.565 |
| Sample008->Sample008 | ML_DOSE | 0.027 | 1.994 | 0.125 | 1.972 | 0.013 | 1.103 | 1.804 | 1.804 | 0.603 |
| Sample009->Sample009 | ML_DOSE | 0.017 | 0.995 | 1.963 | 1.984 | 0.008 | 0.186 | 0.888 | 1.888 | 0.35 |
| Sample010->Sample010 | ML_DOSE | 0.026 | 1.994 | 1.984 | 1.981 | 0.013 | 0.129 | 1.789 | 1.789 | 0.604 |

・hoge.mlinfo上で "A11"=1 となっているSNPを変換対象とし、以下の変換を行う。

hoge.mlinfo上の "A11", "A12"列を入れ替え、"Freq1"を"1-Freq1"に変換する。

hoge.mldose上で "allele dosage" を "2-allele dosage" に変換する。

・その他のMACH出力ファイルのうち、hoge.mlgeno には "Best guess genotype"が記載されている。

・hoge.mlgeno 内のアレル1/2とMACH入力ファイル上のアレル1/2は対応しているため、下記のように "1/1"→0, "1/2"→1, "2/1"→1, "2/2"→2 と変換することにより、hoge_flip.mldose と同じフォーマットのデータを得ることができる。

hoge.mlgenoファイル

```
Sample001->Sample001 ML_GENO 1/1 2/2 2/2 2/2 1/1 1/1 1/2 2/2 1/1
Sample002->Sample002 ML_GENO 1/1 2/2 1/2 2/2 1/1 1/1 2/2 2/2 1/1
Sample003->Sample003 ML_GENO 1/1 2/2 2/1 2/2 1/1 1/1 2/2 2/2 1/1
Sample004->Sample004 ML_GENO 1/1 2/2 2/2 2/2 2/1 1/1 2/2 2/2 1/1
Sample005->Sample005 ML_GENO 1/1 1/2 2/2 2/2 1/1 1/1 2/2 2/2 1/1
Sample006->Sample006 ML_GENO 1/1 2/2 2/2 2/2 1/1 1/1 2/2 2/2 1/1
Sample007->Sample007 ML_GENO 1/1 2/2 2/2 2/2 1/1 1/1 2/2 2/2 1/1
Sample008->Sample008 ML_GENO 1/1 2/2 1/1 2/2 1/1 1/2 2/2 2/2 1/1
Sample009->Sample009 ML_GENO 1/1 1/2 2/2 2/2 1/1 1/1 1/2 2/2 1/1
Sample010->Sample010 ML_GENO 1/1 2/2 2/2 2/2 1/1 1/1 2/2 2/2 1/1
```

↓

hoge_flip.mlgenoファイル

```
Sample001->Sample001 ML_GENO 0 2 2 2 0 0 1 2 0
Sample002->Sample002 ML_GENO 0 2 1 2 0 0 2 2 0
Sample003->Sample003 ML_GENO 0 2 1 2 0 0 2 2 0
Sample004->Sample004 ML_GENO 0 2 2 2 1 0 2 2 0
Sample005->Sample005 ML_GENO 0 1 2 2 0 0 2 2 0
Sample006->Sample006 ML_GENO 0 2 2 2 0 0 2 2 0
Sample007->Sample007 ML_GENO 0 2 2 2 0 0 2 2 0
Sample008->Sample008 ML_GENO 0 2 0 2 0 1 2 2 0
Sample009->Sample009 ML_GENO 0 1 2 2 0 0 1 2 0
Sample010->Sample010 ML_GENO 0 2 2 2 0 0 2 2 0
```

・上記の変換を行うperlスクリプトとして、flipmlgeno_with_mlinfo.pl を作成した。

入力ファイル: hoge.mlinfo
 hoge.mldose
 hoge.mlgeno

コマンド: "perl flipmlgeno_with_mlinfo.pl hoge"

出力ファイル: hoge_flip.mlinfo ... hoge.mlinfo に変換を施行したもの。
 hoge_flip.mldose ... hoge.mldose に変換を施行したもの。
 hoge_flip.mlgeno ... hoge.mlgeno に変換を施行したもの。
 hoge_flip.ped ... hoge_flip.mlgeno をplink形式にしたもの。
 hoge_flip.sampledose.txt ... hoge_flip.mldose のサンプルID。
 hoge_flip.samplegeno.txt ... hoge_flip.mlgeno のサンプルID。

④:【手順2】 mach2datを用いたケースコントロール解析

・mach2datを用いて、allele dosage に対するロジスティック回帰を行う。

入力ファイル: hoge_flip.mlinfo
 hoge_flip.mldose
 Phen.dat ... フェノタイプ名を記載したファイル。"A CaseControl"のように、"A (形質名)"と記載する。
 Phen.ped ... フェノタイプ情報を記載したファイル。形式は下記を参照。

```
Sample001 Sample001 0 0 0 2
Sample002 Sample002 0 0 0 2
Sample003 Sample003 0 0 0 2
Sample004 Sample004 0 0 0 2
Sample005 Sample005 0 0 0 1
Sample006 Sample006 0 0 0 1
Sample007 Sample007 0 0 0 1
Sample008 Sample008 0 0 0 1
Sample009 Sample009 0 0 0 0
Sample010 Sample010 0 0 0 0
```

| | |
|-----|----------------------------------|
| 第1列 | サンプルID (MACH入力ジェノタイプファイルの第1列に同じ) |
| 第2列 | サンプルID (MACH入力ジェノタイプファイルの第2列に同じ) |
| 第3列 | 使用されない |
| 第4列 | |
| 第5列 | |
| 第6列 | |

コマンド: "mach2dat -d Phen.dat -p Phen.ped -i hoge_flip.mlinfo -g hoge_flip.mldose --rsqcutoff -1 > (出力ファイル名)"

hoge_flip.mldose を hoge_flip.mlgeno に変えた場合、Best-guess genotypeに対する解析を行うことが可能。
 結果が標準出力に返されるため、リダイレクト等で結果をファイルに書き出す必要がある。
 Rsq値に関わらず全SNPの解析結果を得るため、"--rsqcutoff -1"としておく。

出力ファイル例:

| ↓(A) | ↓(B) | | | | ↓(C) | | | ↓(D) | | | |
|-------------|----------|---------|--------|--------|---------|--------|--------|-----------|--------|---------|--------|
| TRAIT | MARKER | ALLELES | FREQ1 | RSQR | EFFECT1 | OR | STDERR | WALDCHISQ | PVALUE | LRCHISQ | LRPVAL |
| CaseControl | rs000001 | 2,1 | 0.9956 | 0.0752 | -0.649 | 0.523 | 1.119 | 0.3363 | 0.562 | 0.3294 | 0.566 |
| CaseControl | rs000002 | 2,1 | 0.9792 | 0.0556 | -0.202 | 0.817 | 0.448 | 0.2031 | 0.6523 | 0.2018 | 0.6533 |
| CaseControl | rs000003 | 2,1 | 0.9937 | 0.0033 | 3.413 | 30.356 | 3.279 | 1.0834 | 0.2979 | 1.0961 | 0.2951 |
| CaseControl | rs000004 | 2,1 | 0.0147 | 0.0126 | 0.689 | 1.991 | 1.061 | 0.4214 | 0.5162 | 0.4124 | 0.5208 |
| CaseControl | rs000005 | 2,1 | 0.0057 | 0.005 | -2.084 | 0.124 | 3.202 | 0.4238 | 0.515 | 0.4316 | 0.5112 |
| CaseControl | rs000006 | 2,1 | 0.0737 | 0.0285 | -0.251 | 0.778 | 0.367 | 0.4678 | 0.494 | 0.4718 | 0.4922 |
| CaseControl | rs000007 | 2,1 | 0.0371 | 0.0209 | 0.384 | 1.469 | 0.563 | 0.4657 | 0.495 | 0.4612 | 0.4971 |
| CaseControl | rs000008 | 2,1 | 0.907 | 0.0325 | -0.203 | 0.816 | 0.301 | 0.4546 | 0.5001 | 0.4532 | 0.5008 |
| CaseControl | rs000009 | 2,1 | 0.907 | 0.0332 | -0.198 | 0.82 | 0.297 | 0.4425 | 0.5059 | 0.4412 | 0.5066 |

| | |
|-----|--|
| (A) | Phen.datで指定した形質名 |
| (B) | hoge_flip.mlinfo から抽出されたSNP情報 |
| (C) | アレルA1の effect size, オッズ比 (=exp(effect size)), effect size のSE* |
| (D) | Wald検定、LR検定による χ^2 値及びP値** |

*: 【手順1】でファイル変換実行後、アレルA1=アレル2 となっているため、アレル2のeffect sizeとして扱うことができる。

** : Wald検定による解析結果を採用する方針とする。

⑤:【手順3】 ケース群、コントロール群におけるアレル頻度の抽出

- ・mach2dat はケース群・コントロール群それぞれにおけるアレル頻度を出力しないため、別途集計する必要がある。
- ・アレル頻度集計ツールとして、javaプログラムCalcmlDoseFreq.jar を作成した。

入力ファイル: hoge_flip.mlinfo
 hoge_flip.mldose
 Phen.ped

コマンド: "java -jar CalcmlDoseFreq.jar -prefix hoge_flip -phen Phen.ped"

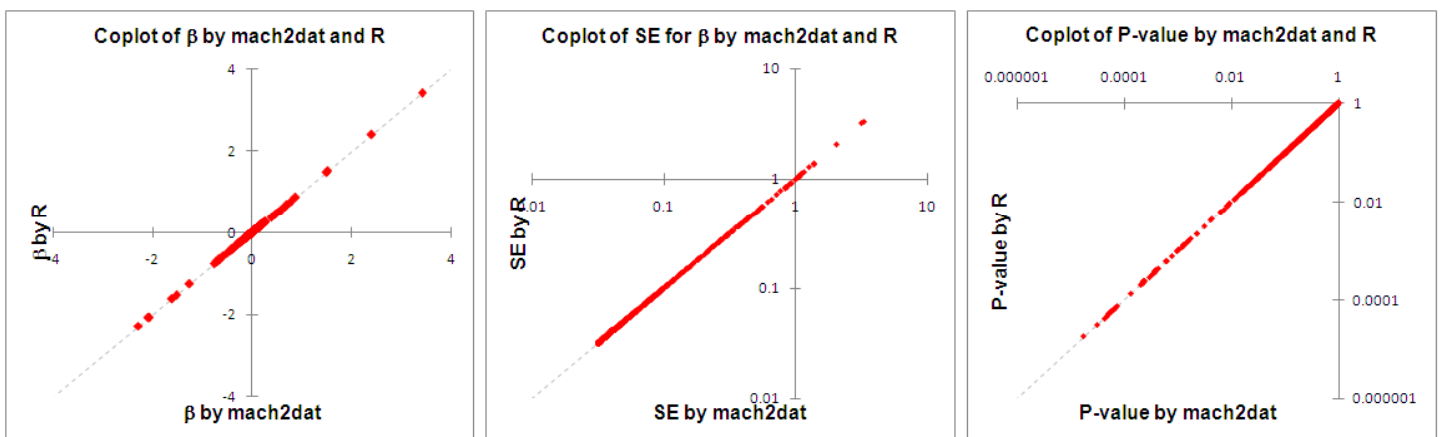
出力ファイル: hoge_flip.freq ... ケース群、コントロール群、データ無し群、全サンプル、のアレル頻度*を記載
 hoge_flip.freq.log ... ログファイル

*: アレル頻度は、(allele dosage のサンプル群内平均値)/2 で定義した。

～参考資料～

○: mach2datによる計算結果の確認

- ・同一データに対し、mach2dat による計算結果と統計ソフトRによる解析結果を比較した。
- ・R version 2.9.0 に実装されたglm関数を用いた。



⇒Beta(=effect size), SE, P値において、結果の一致を確認した。

ご不明な点や、ご質問は [yokada-tky \(at\) umin.ac.jp](mailto:yokada-tky@umin.ac.jp) までご連絡下さい。