

ANALISIS DATA SNP (SINGLE NUCLEOTIDE POLYMORPHISM) DENGAN METODE CHI-SQUARE

Vera Manondang Damaianty Butarbutar¹⁾, Adi Setiawan²⁾, Tundjung Mahatma³⁾

¹Fakultas Sains dan Matematika, Universitas Kristen Satya Wacana
email: 662016012@student.uksw.edu,

²Fakultas Sains dan Matematika, Universitas Kristen Satya Wacana
Email: adi.setiawan@staff.uksw.edu

³Fakultas Sains dan Matematika, Universitas Kristen Satya Wacana
email: tundjung.mahatma@uksw.edu

ABSTRAK

Pada umumnya di antara dua orang secara acak memiliki 99,9% persamaan genetik dan sisanya berbeda atau yang disebut dengan SNP. Penelitian ini mengembangkan penggunaan perhitungan chi-square dalam data genetik SNP (Single Nucleotide Polymorphism atau dikenal dengan Nukleotida Polimorfisme Tunggal) HapMap terhadap ras CEU-Eropa dan ras Yoruba-Afrika. Tujuannya adalah untuk mengidentifikasi apakah terdapat hubungan antara penanda genotip (marker) dan fenotipnya. Untuk mengidentifikasinya, dilakukan perhitungan statistik dengan perhitungan chi-square, kemudian dilakukan penolakan atau penerimaan terhadap hipotesis yang menyatakan ada atau tidaknya hubungan antara penanda genotip dan fenotipnya. Hipotesis ditolak dan diterima bergantung dari taraf signifikansi yang dipakai, dan dalam penelitian ini menggunakan taraf signifikansi sebesar $\alpha = 0.05$, $\alpha = 10^{-15}$ serta beberapa taraf signifikansi lainnya. Perhitungan chi-square nantinya diterapkan ke dalam dua penanda, yaitu penanda genotip dan penanda alel. Diperoleh lebih banyak lokasi SNP yang signifikan pada penanda alel dibanding penanda genotip, yaitu sebesar 3932 SNP untuk penanda genotip, dan sebanyak 5108 SNP untuk penanda alel dengan taraf signifikansi $\alpha = 0.05$. dengan hasil tersebut metode chi-square dapat dipakai untuk membantu menentukan keterkaitan antara penanda genotip dan fenotipnya.

Kata Kunci: HapMap, chi-square, taraf signifikansi, p-value

1. PENDAHULUAN

Sebagian besar materi genetik yang dimiliki manusia adalah sama, dan hanya sedikit perbedaan genetik (Ding & Guo, 2018). Antara dua manusia yang dipilih secara acak, terdapat 99,9% urutan DNA yang identik dan sisanya kurang dari 0.1% mengandung variasi urutan DNA. Variasi urutan ini disebut dengan SNP (Single Nucleotide Polymorphism) atau dikenal dengan Nukleotida Polimorfisme Tunggal yang menimbulkan keragaman dalam populasi. Keragaman tersebut antara lain seperti suatu penyakit turunan, respon terhadap suatu obat, warna kulit, warna mata dan lainnya.

SNP dalam ilmu genetika lebih dikenal untuk mengidentifikasi kerentanan terhadap suatu penyakit kompleks. Studi asosiasi seluruh genom atau dikenal

dengan *Genome-wide association studies* (GWAs) adalah suatu studi genetika yang bertujuan untuk mengidentifikasi suatu varian genetik yang biasa terjadi terhadap resiko penyakit terutama penyakit kompleks (Visscher et al., 2017). Studi paling sederhana yang digunakan untuk mengidentifikasi varian genetik itu adalah studi *case-control*. Tujuan dari studi ini adalah untuk menguji apakah suatu penyakit terkait dengan penanda genotip (dalam hal ini adalah SNPnya) atau tidak ((*Statistics for Biology and Health*) Gang Zheng, Yaning Yang, Xiaofeng Zhu, Robert C, n.d.).

Telah dilakukan beberapa penelitian terkait keterkaitan fenotip dan genotip yang mendasari penelitian ini, antara lain penelitian oleh Voisey dkk, 2010 dengan metode *Pearson's chi square test*

memperoleh beberapa lokasi SNP yang signifikan terkait penyakit Schizophrenia. Selain itu juga penelitian lain yang terkait adalah penelitian oleh Sun dkk, 2019 terkait studi *Genome-wide* terhadap sifat ketergantungan alkohol pada kaum pria suku Han Cina. Dalam penelitian itu, diperoleh beberapa SNP signifikan yang mengakibatkan munculnya ketergantungan terhadap alkohol. Beberapa penelitian tersebut mendorong penggunaan metode *chi-square test* untuk melihat keterkaitan antara penanda genotip dan fenotipnya menggunakan data HapMap.

2. METODE PENELITIAN

Data yang digunakan yaitu data HapMap (*Haplotype Map*) khususnya populasi CEU (*Utah Residence with ancestry from northern and western Europe*) atau ras Eropa dan YRI (*Yoruba trios from Ibadan, Nigeria*) atau ras Afrika. Data HapMap berukuran 120×9305 . Data HapMap sudah tersedia dalam paket *SNPassoc* pada aplikasi R.

Data tersebut kemudian dihitung nilai statistiknya dengan uji *chi-square*. Untuk mengidentifikasi ada atau tidaknya hubungan antara suatu penanda genotip dengan penanda fenotipnya, digunakan perhitungan statistik menggunakan metode *chi-square*. Dalam menentukan ada atau tidaknya hubungan tersebut, dilakukan perhitungan *p-value* dengan metode *chi-square* untuk berikutnya diambil langkah pengambilan keputusan apakah terdapat hubungan atau tidak. Uji keterkaitan antara genotip dan fenotip diuji dengan hipotesis :

H₀ : tidak ada hubungan antara penanda genotip dan fenotipnya,

H₁ : ada hubungan antara penanda genotip dan fenotipnya.

Pada pengolahan data ini, dibagi ke dalam dua cara yaitu perhitungan *chi-square* dengan penanda tunggal berbasis genotip (*single marker genotype-based*).

1.1. Chi-square test

Uji statistik *chi-square* (Q^2) sebagai berikut, lihat (Setiawan, 2007):

$$Q^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

dengan,

O_{ij} = banyaknya hasil yang diobservasi pada baris ke- i dan kolom ke- j ,

E_{ij} = banyaknya jumlah yang diharapkan masuk dalam sel (i, j) dimana $i = 1, 2$ dan $j = 1, 2, 3$.

1.2. Penanda tunggal berbasis genotip

Diambil sampel random dari n *case* dan m *control*. Adapun untuk pemisahan penanda genotip *case* dan *control* dapat dinyatakan pada Tabel 1. dengan $\mathbf{X} = (X_1, X_2, X_3)$ dan $\mathbf{Y} = (Y_1, Y_2, Y_3)$ adalah jumlah dari genotip AA, Aa dan aa dalam sampel *case* dan *control* berturut-turut (Setiawan, 2007). Untuk variabel N adalah jumlah penanda genotip untuk fenotip *case* (populasi CEU menjadi variabel *case*), variabel M adalah jumlah penanda genotip untuk fenotip *control* (populasi YRI menjadi variabel *control*), dan M+N adalah jumlah keseluruhan penanda genotip untuk fenotip *case* dan *control*.

Tabel 1. Jumlah penanda genotip untuk *case* dan *control*

| | AA | Aa | aa | Jumlah |
|--------|-------------|-------------|-------------|--------|
| CEU | X_1 | X_2 | X_3 | N |
| YRI | Y_1 | Y_2 | Y_3 | M |
| Pooled | $X_1 + Y_1$ | $X_2 + Y_2$ | $X_3 + Y_3$ | N+M |

1.3. Penanda tunggal berbasis alel

Dapat dilihat pada Tabel 2, vektor \mathbf{X} adalah jumlah alel A penanda *case* dan vektor \mathbf{Y} adalah jumlah alel A penanda *control*. Sedangkan $2n - X$ adalah jumlah dari alel a penanda *case* dan $2m - Y$ adalah jumlah alel a penanda *control* dalam sampel *control*

dan *case* berturut-turut (Setiawan, 2007).

Tabel 2. Jumlah penanda alel untuk *case* dan *control*

| | <i>A</i> | <i>a</i> | Jumlah |
|--------|----------|--------------|------------|
| CEU | <i>X</i> | $2n - X$ | $2n$ |
| YRI | <i>Y</i> | $2m - Y$ | $2m$ |
| Pooled | $X + Y$ | $2(n + m) -$ | $2(n + m)$ |

Statistik Q^2 untuk penanda berbasis genotip berdistribusi *chi-square* derajat bebas sebesar 2, dan untuk penanda berbasis alel berderajat bebas sebesar 1. Syarat penolakan dan penerimaan hipotesis adalah sebagai berikut :

- $Q_{hitung}^2 \leq Q_{tabel}^2$ dan *p-value* > α maka H_0 diterima
- $Q_{hitung}^2 \geq Q_{tabel}^2$ dan *p-value* < α maka H_0 ditolak.

Metode penanda tunggal berbasis genotip di penelitian ini menggunakan taraf signifikansi α sebesar 0,05 berderajat bebas dua dan nilai Q^2 adalah sebesar 5,991. Sedangkan untuk metode penanda tunggal berbasis alel menggunakan taraf signifikansi α sebesar 0,05 nilai Q^2 nya adalah sebesar 3,841.

Taraf signifikansi terkoreksi Bonferonni dan taraf signifikansi lainnya.

Koreksi Bonferroni bertujuan untuk mengontrol kemungkinan muncul setidaknya satu tanda positif yang keliru (salah) dengan menghitung batas *p-value*. Koreksi Bonferroni menyesuaikan *p-value* dimana dilakukan evaluasi terhadap taraf signifikansi berdasarkan jumlah total tes yang dilakukan (Gelman, Hill, & Yajima, 2012). Untuk menyesuaikan *p-value* digunakan koreksi Bonferroni dengan rumus sebagai berikut :

$$\alpha = \frac{\pi}{B}$$

dengan,

π = banyaknya hipotesis,

B = jumlah penanda yang dievaluasi.

Sehingga taraf signifikansi α yang baru untuk data HapMap yang dipakai pada

penelitian ini adalah sebesar 5×10^{-4} yang berarti akan dikatakan signifikan apabila *p-value* di bawah 5×10^{-4} .

Untuk menangani kekhawatiran pada beberapa pengujian terhadap data HapMap, sering digunakan taraf signifikansi α sebesar 5×10^{-8} di sebagian besar studi GWA (Fadista, Manning, Florez, & Groop, 2016). Selain menggunakan taraf signifikansi sebesar 0.05, 5×10^{-6} dan 5×10^{-8} , untuk populasi HapMap dengan metode genotip dan alel pada penelitian ini juga menggunakan taraf signifikansi sebesar 5×10^{-15} sebagai ambang batas untuk mengontrol kesalahan Tipe 1 seluruh penanda genotip. Sebagai catatan, taraf signifikansi 10^{-15} adalah nilai taraf signifikansi *default* yang terdapat pada program R.

3. HASIL DAN PEMBAHASAN

1. Pengolahan data:

Langkah-langkah yang digunakan yaitu:

- Mengelompokkan data European (CEU) dan Yoruba (YRI).
- Dari kelompok data European (CEU) dan Yoruba (YRI), dipisahkan ke dalam tabel 2×3 untuk metode penanda tunggal berbasis genotip dan dari tabel 2×3 itu, dipisahkan lagi ke dalam tabel 2×2 kemudian dilanjutkan ke perhitungan dengan metode penanda tunggal berbasis alel.
- Menentukan hipotesis.
- Menentukan taraf signifikansi. Ada beberapa taraf signifikansi yang dipakai antara lain :
 - $\alpha = 0,05$ (taraf signifikansi yang sering dipakai dalam metode *chi-square* test
 - 5×10^{-6} (taraf signifikansi terkoreksi Bonferroni)
 - 5×10^{-8} (taraf signifikansi yang umum digunakan pada GWAs, dan

- d. 5×10^{-15} (taraf signifikansi default aplikasi R).
 - 5) Menghitung nilai statistik dengan metode *chi-square test*
 - 6) Pengambilan keputusan
2. Hasil analisis awal:
- 2.1. Metode penanda tunggal berbasis genotip

Tabel 3. Jumlah penanda genotip CEU dan YRI untuk sampel rs11260616

| | A/A | A/T | T/T | Jumlah |
|--------|-----|-----|-----|--------|
| CEU | 28 | 31 | 1 | 60 |
| YRI | 37 | 19 | 4 | 60 |
| Pooled | 65 | 50 | 5 | 120 |

Berdasarkan data HapMap yang diperoleh dalam paket program R yaitu *SNPassoc*, sebagai sampel diperhatikan lokasi SNP yaitu rs11260616. Untuk sampel lokasi rs11260616 dengan uji *chi-square* diperoleh nilai statistik Q^2_{hitung} adalah 5,92. Nilai Q^2_{hitung} dipakai untuk menghitung *p-value*. Selanjutnya diperoleh *p-value* untuk lokasi SNP rs11260616 sebesar 0,052. Dari hasil perhitungan *chi-square*, diperoleh bahwa hasil *chi-square* hitung lebih kecil dibanding *chi-square* dari tabel. Dari pengujian dengan menggunakan taraf signifikansi $\alpha=0,05$, *p-value* dari Q^2 adalah 0,052 sehingga H_0 diterima dan dapat disimpulkan bahwa tidak ada hubungan antara penanda genotip dan fenotipnya.

Selanjutnya dilakukan langkah yang sama terhadap lokasi SNP lainnya. Dengan *coding* di R, diperoleh jumlah lokasi SNP yang signifikan sebanyak 3932 lokasi yang signifikan (*p-value* kurang dari $\alpha = 0,05$). Pada perolehan nilai signifikansi, terdapat *p-value* dengan rentang nilai antara 0 hingga 1. *P-value* mendekati nol (atau secara praktis bernilai 0) diperoleh dari nilai Q^2 yang sangat besar. Sampel lima nilai statistik Q^2 yang terbesar dapat dilihat pada Tabel 4.

Tabel 4. Nilai Q^2 untuk *p-value* yang bernilai 0 beserta lokasi SNPnya

| No | Id SNP | Nilai Q^2 | Lokasi SNP |
|----|------------|-------------|----------------|
| 1 | rs10868791 | 112.3524 | chr9:88539211 |
| 2 | rs2370893 | 112.2580 | hr14:78997500 |
| 3 | rs7851392 | 101.5385 | chr9:137820446 |
| 4 | rs6814827 | 101.3483 | chr4: 90419517 |
| 5 | rs10805068 | 100.5372 | chr4:73756570 |

Tabel 5. Sampel 10 Lokasi SNP pada data HapMap yang signifikan menggunakan metode genotip

| No | Id Lokasi SNP | Lokasi SNP | <i>p-value</i> |
|----|---------------|----------------|----------------------------|
| 1 | rs3124625 | chr1:11440415 | $1,110223 \times 10^{-16}$ |
| 2 | rs2189174 | chr4:112362455 | $2,220446 \times 10^{-16}$ |
| 3 | rs6452430 | chr5:81827755 | $3,330669 \times 10^{-16}$ |
| 4 | rs4952644 | chr2:40721629 | $4,440892 \times 10^{-16}$ |
| 5 | rs3792076 | chr2:230792323 | $5,551115 \times 10^{-16}$ |
| 6 | rs3901704 | chr3:171505620 | $6,661338 \times 10^{-16}$ |
| 7 | rs2258475 | chr2:130410315 | $7,771561 \times 10^{-16}$ |
| 8 | rs7803071 | chr7:83830592 | $8,881784 \times 10^{-16}$ |
| 9 | rs7163946 | chr15:76473708 | $9,992007 \times 10^{-16}$ |
| 10 | rs6880750 | chr5:117221040 | $1,110223 \times 10^{-15}$ |

Suatu lokasi SNP dikatakan signifikan atau membawa pengaruh terhadap genotip atau sifat suatu ras apabila *p-value* bernilai lebih kecil daripada 0,05 seperti yang dinyatakan pada Tabel 5. Terdapat juga beberapa lokasi SNP dengan *p-value* sebesar 1,00 yang berarti lokasi SNP tersebut tidak signifikan atau tidak menentukan munculnya suatu sifat dari ras CEU ataupun YRI. *P-value* sebesar 1,00 terjadi karena lokasi SNP itu memiliki penanda genotip yang sama dan nilai Q^2 mendekati nol (atau secara praktis bernilai 0). Dikatakan SNP apabila dalam suatu lokasi, penanda genotipnya berbeda. Terdapat sebanyak 3932 lokasi SNP yang

signifikan berdasarkan metode penanda genotip.

2.2. Metode penanda tunggal berbasis alel

Tabel 6. Jumlah penanda alel CEU dan YRI untuk sampel rs11260616

| | A | T | Jumlah |
|--------|-----|----|--------|
| CEU | 87 | 33 | 120 |
| YRI | 93 | 27 | 120 |
| Pooled | 180 | 60 | 240 |

Untuk penanda tunggal, kita hanya memerlukan penanda A dan T saja. Dari hasil pengelompokan alel CEU dan YRI pada Tabel 6, diperoleh nilai statistik Q^2_{hitung} sebesar 0,8. Nilai Q^2_{hitung} kemudian dipakai untuk menghitung *p-value*. Selanjutnya diperoleh *p-value* untuk lokasi SNP rs11260616 sebesar 0,37 (dengan bantuan R). Dari hasil perhitungan *chi-square* diperoleh bahwa hasil *chi-square* hitung lebih kecil dibanding *chi-square* dari tabel (atau yang dihitung menggunakan R), dan *p-value* lebih besar dibanding taraf signifikansi $\alpha=0,05$. Sehingga H_0 diterima dan dapat disimpulkan bahwa tidak ada hubungan antara penanda genotip dan fenotip.

Selanjutnya dilakukan langkah yang sama terhadap lokasi SNP lainnya. Dengan *coding* di R, diperoleh jumlah lokasi SNP yang signifikan adalah sebanyak 4220 lokasi yang signifikan (*p-value* kurang dari $\alpha = 0,05$).

Tabel 7. Nilai Q^2 untuk *p-value* yang (mendekati nol) bernilai 0 beserta lokasi SNPnya.

| No | Id Lokasi SNP | Lokasi SNP | Nilai Q^2 |
|----|---------------|--------------------|-------------|
| 1 | rs10868791 | chr9:8853921 1 | 182,7542 |
| 2 | rs9909962 | chr17:534378 73 | 169,4915 |
| 3 | rs6670842 | chr1:5304726 9 | 166,7129 |
| 4 | rs2370893 | chr14:789975 00 | 165,6338 |
| 5 | rs6814827 | chr4:9041951 7 | 163,9077 |

Tabel 11. Sampel 10 Lokasi SNP pada data HapMap yang signifikan menggunakan metode alel

| No | Id Lokasi SNP | Lokasi SNP | p-value |
|----|----------------|---------------------|----------------------------|
| 1 | rs3213564 | chr1:1723 80781 | $1,110223 \times 10^{-16}$ |
| 2 | rs261631 | chr5:1698 30772 | $2,220446 \times 10^{-16}$ |
| 3 | rs1272760 5 | chr1:3051 4397 | $3,330669 \times 10^{-16}$ |
| 4 | rs9882569 | chr3:3544 2918 | $4,440892 \times 10^{-16}$ |
| 5 | rs6489167 | chr12:127 304927 | $5,551115 \times 10^{-16}$ |
| 6 | rs7019462 | chr9:8744 4328 | $6,661338 \times 10^{-16}$ |
| 7 | rs1578051 | chr13:551 57126 | $7,771561 \times 10^{-16}$ |
| 8 | rs1445670 | chr1:1888 86748 | $8,881784 \times 10^{-16}$ |
| 9 | rs4926966 | Chr1:4769 8950 | $9,992007 \times 10^{-16}$ |
| 10 | rs6766679 | Chr3:1167 67947 | $1,110223 \times 10^{-15}$ |

Sama halnya dengan metode penanda genotip, suatu lokasi SNP dikatakan signifikan atau membawa pengaruh terhadap genotip atau sifat suatu ras apabila *p-value* bernilai lebih kecil daripada 0,05 seperti yang dinyatakan pada Tabel 11. Terdapat juga beberapa lokasi SNP dengan *p-value* sebesar 1,00 yang berarti lokasi SNP tersebut tidak signifikan atau tidak menentukan munculnya suatu sifat dari ras CEU ataupun YRI. *P-value* sebesar 1,00 terjadi karena lokasi SNP itu memiliki penanda genotip yang sama dan nilai Q^2 mendekati nol (atau secara praktis bernilai 0). Dikatakan SNP apabila dalam suatu lokasi, penanda genotipnya berbeda. Terdapat sebanyak 3932 lokasi SNP yang signifikan berdasarkan metode penanda alel.

2.3. Menghitung jumlah SNP signifikan dengan taraf signifikansi yang terkoreksi Bonferroni dan taraf signifikansi lainnya

Berikut taraf signifikansi terkoreksi Bonferroni :

$$\alpha = \frac{\pi}{B} = \frac{0.05}{9308} = 5,371723 \times 10^{-6}$$

Batas taraf signifikan terkoreksi Bonferonni pada penelitian ini adalah sebesar 5.371723×10^{-6} . Untuk metode penanda genotip, diperoleh sebanyak 2346 lokasi yang dianggap signifikan berdasarkan taraf signifikan terkoreksi Bonferonni. Untuk metode penanda tunggal, diperoleh sebanyak 2093 lokasi yang dianggap signifikan berdasarkan taraf signifikansi terkoreksi Bonferonni.

Dalam studi *Genome wide Association*, untuk studi tentang populasi Eropa biasanya digunakan batas nilai signifikansi sebesar 5×10^{-8} . Pada metode berbasis penanda genotip terdapat 1031 lokasi yang signifikan sedangkan pada metode berbasis penanda alel terdapat 1371 lokasi yang signifikan berdasarkan taraf signifikansi yang ditentukan dari *genome wide association studies*. Dalam paket *SNPassoc* di R, ada nilai *default* atau nilai yang sudah ditentukan untuk taraf signifikansi, yaitu sebesar 10^{-15} . Pada metode berbasis penanda genotip terdapat 131 lokasi yang signifikan sedangkan pada metode berbasis penanda alel terdapat 313 lokasi yang signifikan berdasarkan taraf signifikansi yang ditentukan dari nilai default program R.

Pada penelitian lain yang dilakukan oleh Sun dkk. 2019, diperoleh beberapa SNP signifikan yang terkait dengan munculnya ketergantungan terhadap alkohol. Dalam penelitian itu, diperoleh nilai signifikansi berbeda-beda. Pada nilai signifikansi sebesar $6,64 \times 10^{-16}$ diperoleh penanda genotip SNP rs 1229984*ADH1B yang menjadi penyebab ketergantungan tersebut. Sedangkan pada penelitian lain tentang keterkaitan suatu penanda genotip dengan penyebab penyakit *Schizophrenia* yang dilakukan oleh Voisey dkk. 2010 diperoleh 13 SNP yang signifikan terkait penyakit tersebut pada taraf signifikansi

0.05. Ke-13 lokasi SNP tersebut yaitu rs9476886, rs1997679, rs2743857, rs3829893, rs7758659, rs9370823, rs9370822, rs4236167, rs16876589, rs4712253, rs742106, rs17470454, rs1047631.

Apabila dibandingkan dengan penelitian ini, jumlah lokasi SNP yang signifikan sangatlah berbeda jauh. Karena pada penelitian ini, penentu sifat dari suatu ras sangat banyak dibandingkan dengan penelitian terhadap suatu sifat/penyakit tertentu seperti yang dilakukan oleh Sun dkk. 2019 juga yang dilakukan oleh Voisey dkk. 2010. Pada penelitian ini dapat dilihat jumlah SNP yang signifikan terbilang sangat banyak, bahkan ketika nilai signifikan semakin kecil masih terdapat ratusan SNP yang signifikan. Hal ini terjadi lantaran penentu suatu ras terdiri dari banyak hal antara lain warna kulit, warna mata, bentuk rambut, postur tubuh dan masih banyak lagi tanda-tanda fenotip yang terlihat. Dengan demikian diasumsikan bahwa jumlah SNP yang diperoleh banyak dapat dikatakan wajar.

4. KESIMPULAN

Berdasarkan pada bab-bab sebelumnya dapat diambil kesimpulan bahwa metode *chi-square* dapat digunakan untuk menentukan *p-value* dari statistik uji dengan prosedur yang telah dijabarkan. Metode ini juga dapat digunakan untuk memperoleh *p-value* suatu lokasi SNP yang menentukan apakah lokasi tersebut menjadi penentu munculnya sifat fenotip berdasarkan beberapa taraf signifikansi yang sudah ditentukan.

Dalam menghitung jumlah lokasi SNP yang signifikan, diperoleh jumlah pada metode genotip lebih sedikit dibandingkan dengan metode fenotip. Namun semua lokasi SNP yang signifikan dengan metode genotip juga signifikan pada metode penanda alel.

5. REFERENSI

- (*Statistics for Biology and Health*) Gang Zheng, Yaning Yang, Xiaofeng Zhu, Robert C. (n.d.).
- Ding, X., & Guo, X. (2018). A Survey of SNP Data Analysis. *Big Data Mining and Analytics*, 1(3), 173–190. <https://doi.org/10.26599/bdma.2018.9020015>
- Fadista, J., Manning, A. K., Florez, J. C., & Groop, L. (2016). The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *European Journal of Human Genetics*, 24(8), 1202–1205. <https://doi.org/10.1038/ejhg.2015.269>
- Gelman, A., Hill, J., & Yajima, M. (2012). Why We (Usually) Don't Have to Worry About Multiple Comparisons. *Journal of Research on Educational Effectiveness*, 5(2), 189–211. <https://doi.org/10.1080/19345747.2011.618213>
- Setiawan, A., & A. (2007). *Statistical Analysis of Genetic Data in Twin Studies and Association Studies*. Retrieved from <http://dare.uvu.vu.nl/handle/1871/10741>
- Sun, Y., Chang, S., Wang, F., Sun, H., Ni, Z., Yue, W., ... Shi, J. (2019). Genome-wide association study of alcohol dependence in male Han Chinese and cross-ethnic polygenic risk score comparison. *Translational Psychiatry*, 9(1). <https://doi.org/10.1038/s41398-019-0586-3>
- Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics*, 101(1), 5–22. <https://doi.org/10.1016/j.ajhg.2017.06.005>
- Voisey, J., Swagell, C. D., Hughes, I. P., Lawford, B. R., Young, R. M., & Morris, C. P. (2010). Analysis of HapMap tag-SNPs in dysbindin (DTNBP1) reveals evidence of consistent association with schizophrenia. *European Psychiatry*, 25(6), 314–319. <https://doi.org/10.1016/j.eurpsy.2009.11.011>