

## BAB V

### HASIL DAN PEMBAHASAN

#### 5.1 Tahap analisis

Tahapan yang perlu dilakukan sebelum analisis yang dilakukan oleh penulis, yaitu menginput atau membaca data dalam program R, setelah data terbaca/terinput langkah selanjutnya adalah melakukan instalasi packages yang digunakan dalam analisis. Setelah packages berhasil diinstal langkah selanjutnya adalah memanggil library packages yang sebelumnya sudah diinstall pada program R.

Setelah data dibaca dalam software R, langkah selanjutnya adalah dengan melihat dari pheno data dengan perintah *ReadAffy*, kemudian melihat karakteristik yang terdapat dalam pheno data, ketika karakteristik dari pheno data didapatkan selanjutnya adalah menampilkan hasil deskriptif dari pheno data. Sehingga diketahui data memiliki dimensi 54 x 54675.

Data berdimensi 54 x 54675 dianalisis melalui beberapa tahapan, *Preprocessing* data merupakan tahapan selanjutnya, beberapa hal yang dilakukan adalah untuk membuang/menghilangkan *background noise* yang tidak sesuai dengan penelitian serta menyesuaikan *cross hibridasi*, menghilangkan variasi non biologis, serta menggabungkan intensitas *probe*. Setelah melakukan *preprocessing* ada data yang dimiliki tidak terjadi pengurangan pada dimensi data. Hal ini dikarenakan data yang terdapat pada *National Center for Biotechnology Information* (GSE 20489) sebagian besar bukan merupakan data mentah namun data yang sudah di *preprocessing*. Sehingga dimensi data tetap pada dimensi 54 x 54675.

Tahapan *preprocessing* merupakan tahapan sebelum melakukan *filtering* dimana sebelum melakukan *filtering* data yang dianalisis harus terbebas dari factor factor yang umumnya terdapat dalam data *bioinformatic* (data gen) beberapa diantaranya seperti *background noise*, *cross hibridasi*, variasi non biologis. Setelah *preprocessing* dilakukan maka tahapan selanjutnya adalah melakukan *filtering* dengan non spesifik *filter*. kegiatan *filtering* adalah untuk mengurangi data yang

memiliki signifikansi yang tidak terlalu berpengaruh dalam analisis data nantinya. Sehingga ketika hasil didapatkan akan menjadi lebih relevan. Pada tahap filtering data didapati pengurangan dimensi pada data dari 54 x 54675 menjadi 54 x 201. Sehingga hasil akhir dari dimensi data sebelum masuk analisis *biclustering* pada tahap filtering berdimensi 54 x 201.

**Tabel 5.1** Hasil Perubahan Dimensi data

	Dimensi Sebelum	Dimensi Sesudah
<i>Preprocessing</i>	54 x 54675	54 x 54675
<i>Filtering</i>	54 x 54675	54 x 201

## 5.2 Ringkasan dan Struktur Data

Pada penelitian yang dilakukan penulis, data yang diperoleh berupa gen series GSE20489 kemudian data tersebut dibaca menggunakan perintah *AffyData*, karena data pada gen series berbentuk cel maka digunakan perintah *AffyData*. Kemudian perintah *ReadAffy* adalah perintah untuk membaca definisi isi dari data, seperti *annotation*, jumlah gen dan jumlah sampel gen. *Pheno* data adalah proses untuk mengetahui informasi dalam series data yang digunakan. Informasi dari data series GSE20489 berupa sejumlah variabel yang digunakan untuk mengetahui karakteristik, dan keterangan-keterangan lainnya.

**Tabel 5.2** Informasi Data

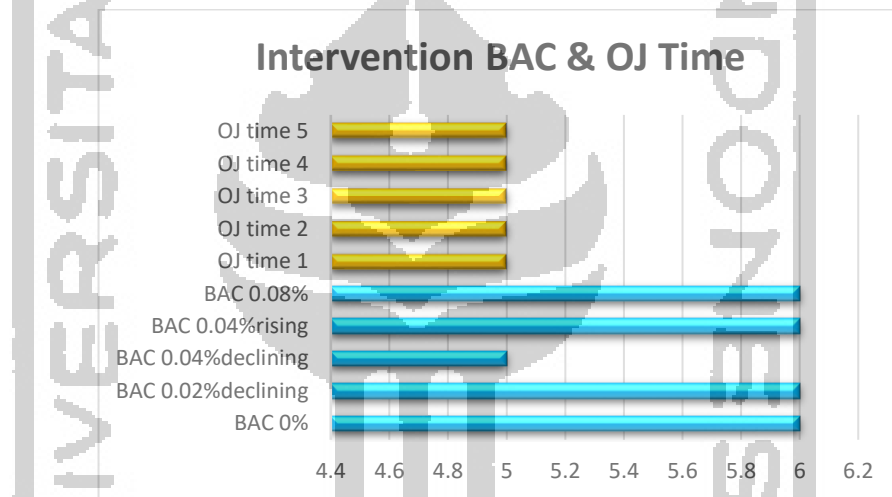
Informasi	Keterangan
Jumlah Sampel	54
Jumlah Gen ( <i>features</i> )	54678
Annotation	hgu133plus2
Tipe	Data frame

Pada **Tabel 5.2** merupakan informasi data, *number of sample* merupakan keterangan dari jumlah orang yang diambil sampel gen darahnya kemudian akan diteliti, tipe data merupakan data frame. *Annotation* atau anotasi DNA adalah proses

mengidentifikasi lokasi gen dan semua wilayah pengkodean yang ada didalam genom dan menentukan apa yang dilakukan oleh gen tersebut. dalam hal ini annotation yang digunakan adalah hgu133plus2.

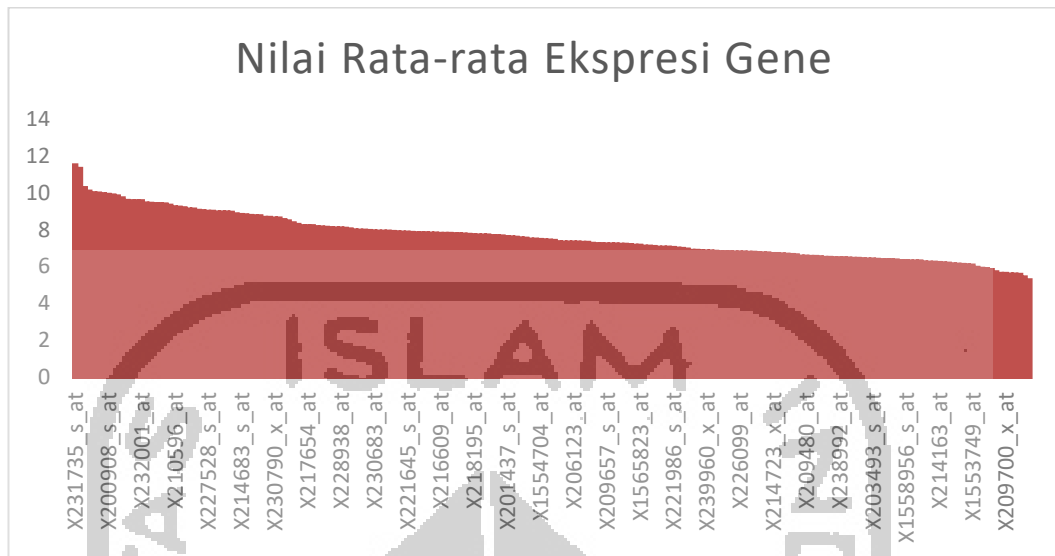
### 5.3 Analisis Deskriptif

Analisis deskriptif adalah analisis yang dilakukan untuk menilai karakteristik dari sebuah sampel atau gambaran umum tentang suatu data yang akan digunakan. Penelitian ini menggunakan grafik yang berupa *pie chart* untuk melihat gambaran variabel data yang digunakan.



**Gambar 5.1** Bar chart Kondisi BAC dan OJ Time

Pada **Gambar 5.1** terdapat kondisi nilai BAC( kadar *alcohol* dalam darah) sample yang dianalisis sebanyak 54 terdiri 5 pasien dari data pada kondisi 0,04% declining, 6 pasien dari data pada kondisi 0,04% rising, sebanyak 6 pasien dari data pada kondisi 0,08%, sebanyak 6 pasien dari data pada kondisi 0,02% declining dan sebanyak 6 pasien pada kondisi 0%. Sebanyak 5 pasien pada kondisi OJ time 1, Sebanyak 5 pasien pada kondisi OJ time 2, Sebanyak 5 pasien pada kondisi OJ time 3, Sebanyak 5 pasien pada kondisi OJ time 4, Sebanyak 5 pasien pada kondisi OJ time 5.



**Gambar 5.2** Nilai Rata-rata Ekspresi Gene

Pada **Gambar 5.2** diatas merupakan nilai rata rata keseluruhan ekspresi gene sebanyak 201 gene dari 54 sample yang diperoleh terdiri dari beberapa kondisi BAC yang sudah disebutkan pada pembahasan sebelumnya. Nilai ekspresi gen yang paling besar ditunjukkan oleh gene X231735\_s\_at dengan nilai sebesar 11.4642908164074 sedangkan untuk nilai ekspresi gene yang paling kecil ditunjukkan oleh gene X1555884\_at.

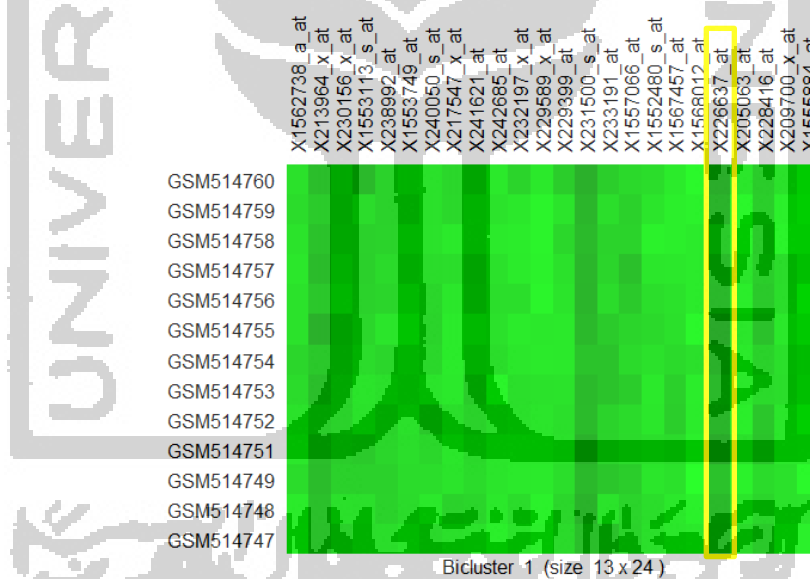
#### 5.4 Analisis Biclustering

Pengelompokan yang dilakukan pada data yang terdiri dari 201 gen dan 54 sample pada metode *Spectral biclustering*, dengan menggunakan *packages biclust* yang berfungsi menyediakan *algoritma Spectral biclustering* pada *software R* yang digunakan.. Hasil pengelompokan data dihasilkan 5 *bicluster* yang terdapat pada Lampiran 2. Hasil dari 5 *bicluster* terbesar ditampilkan pada Tabel 5.3.

**Tabel 5.3** hasil *bicluster*

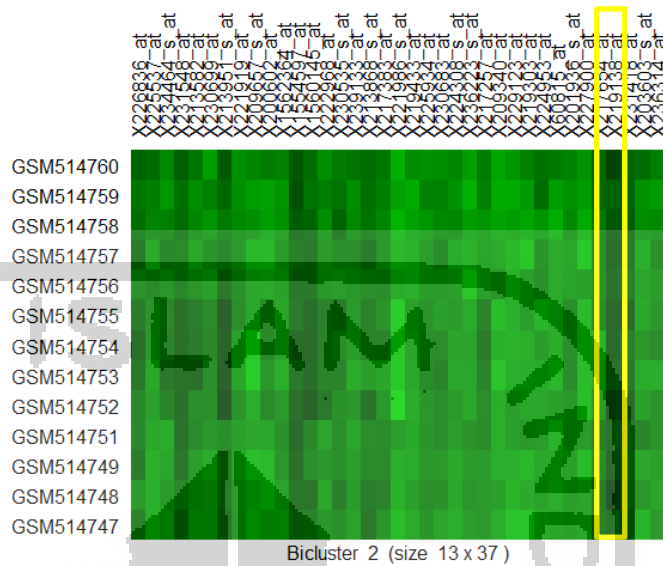
	BC 1	BC 2	BC 3	BC 4	BC 5
Sample	13	13	13	13	13
Gen	24	37	45	35	51

Hasil analisis *bicluster* algoritma *plaid* terbentuk sebanyak 5 *bicluster*, dengan dimensi masing masing *bicluster* yang berbeda-beda. Dimensi *bicluster1* memiliki baris berjumlah 13 dimana menyatakan bayaknya gen dan 24 kolom dimana menyatakan sebuah sampel. Dimensi *bicluster2* memiliki baris berjumlah 13 dimana menyatakan bayaknya gen dan 37 kolom dimana menyatakan sebuah sampel. Dimensi *bicluster3* memiliki baris berjumlah 13 dimana menyatakan bayaknya gen dan 45 kolom dimana menyatakan sebuah sampel. Dimensi *bicluster4* memiliki baris berjumlah 13 dimana menyatakan bayaknya gen dan 35 kolom dimana menyatakan sebuah sampel. Dimensi *bicluster5* memiliki baris berjumlah 13 dimana menyatakan bayaknya gen dan 51 kolom dimana menyatakan sebuah sampel. Pada **Table 5.3** dimensi yang terbesar ada pada *bicluster 5* yaitu terdapat 13 sampel dan 51 gen, sedangkan hasil dimensi *bicluster* terkecil terdapat pada *bicluster1* yaitu 13 sampel dengan 24 gen.



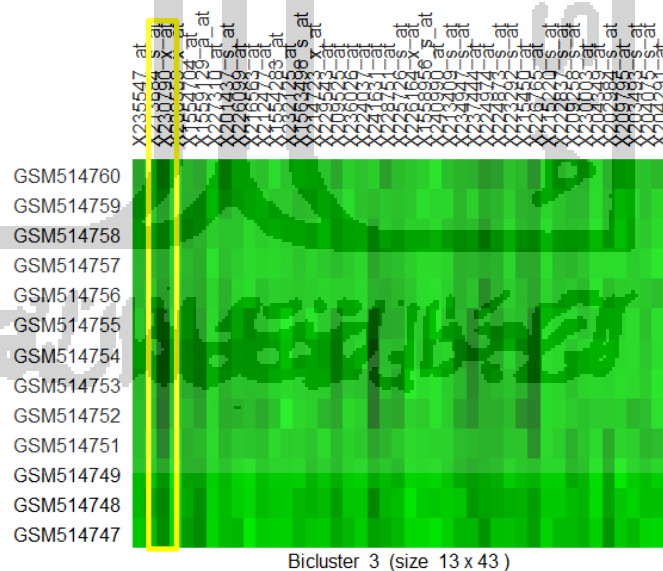
**Gambar 5.3** Heatmap Biclustering 1

Pada Berdasarkan **Gambar 5.3** diperoleh informasi bahwa sebagian besar organ memiliki ekspresi gen relatif sama. Sedangkan gene X231500\_s\_at dan X226637\_at memiliki tingkat ekspresi gen yang tinggi. Hal tersebut di tunjukkan pada daerah didalam persegi Panjang berwarna kuning dengan warna sedikit gelap pada *heatmap*.



**Gambar 5.4** Heatmap Biclustering 2

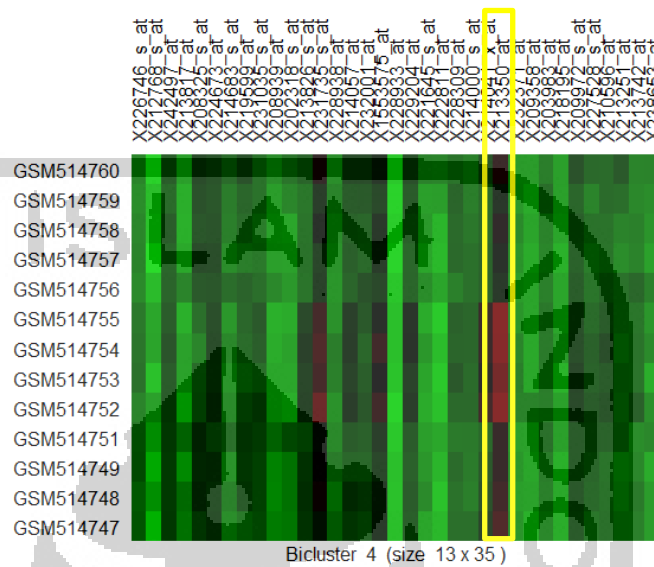
Pada Berdasarkan **Gambar 5.4** diperoleh informasi bahwa tampak beberapa gen memiliki nilai ekspresi gen yang tinggi di banding gen lain pada *biclusternya*. Hal ni di tunjukan pada daerah didalam persegi Panjang berwarna kuning dengan warna pada heatmap yang sebagian gelap. Nilai ekspresi gen tertinggi yaitu X219138\_at sedangkan gen dengan nilai ekspresi terendah yaitu X221986\_s\_at.



**Gambar 5.5** Heatmap Biclustering 3

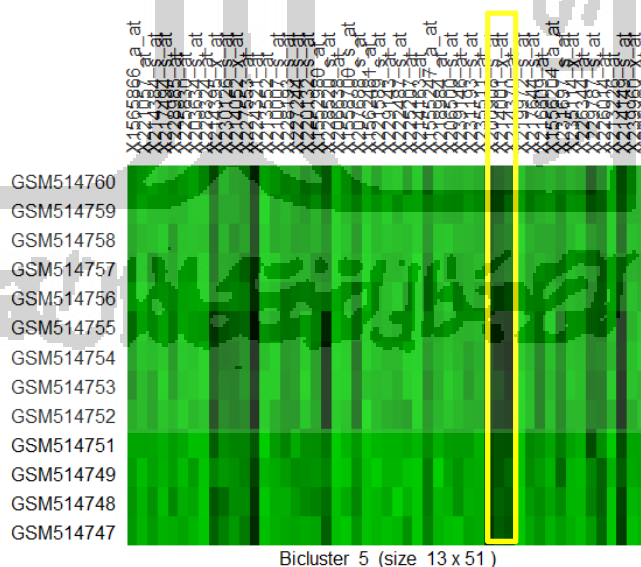
Pada Berdasarkan **Gambar 5.5** diketahui bahwa sebagian besar warna pada heatmap tidak menunjukkan perbedaan antara gen tertentu dengan gen lainnya yang

masih berada pada *biclusternya*. Sedangkan gen X230790\_x\_at merupakan gen yang memiliki nilai ekspresi yang tinggi pada *bilcluster3* yaitu bernilai 5.813.



**Gambar 5.6** Heatmap Biclustering 4

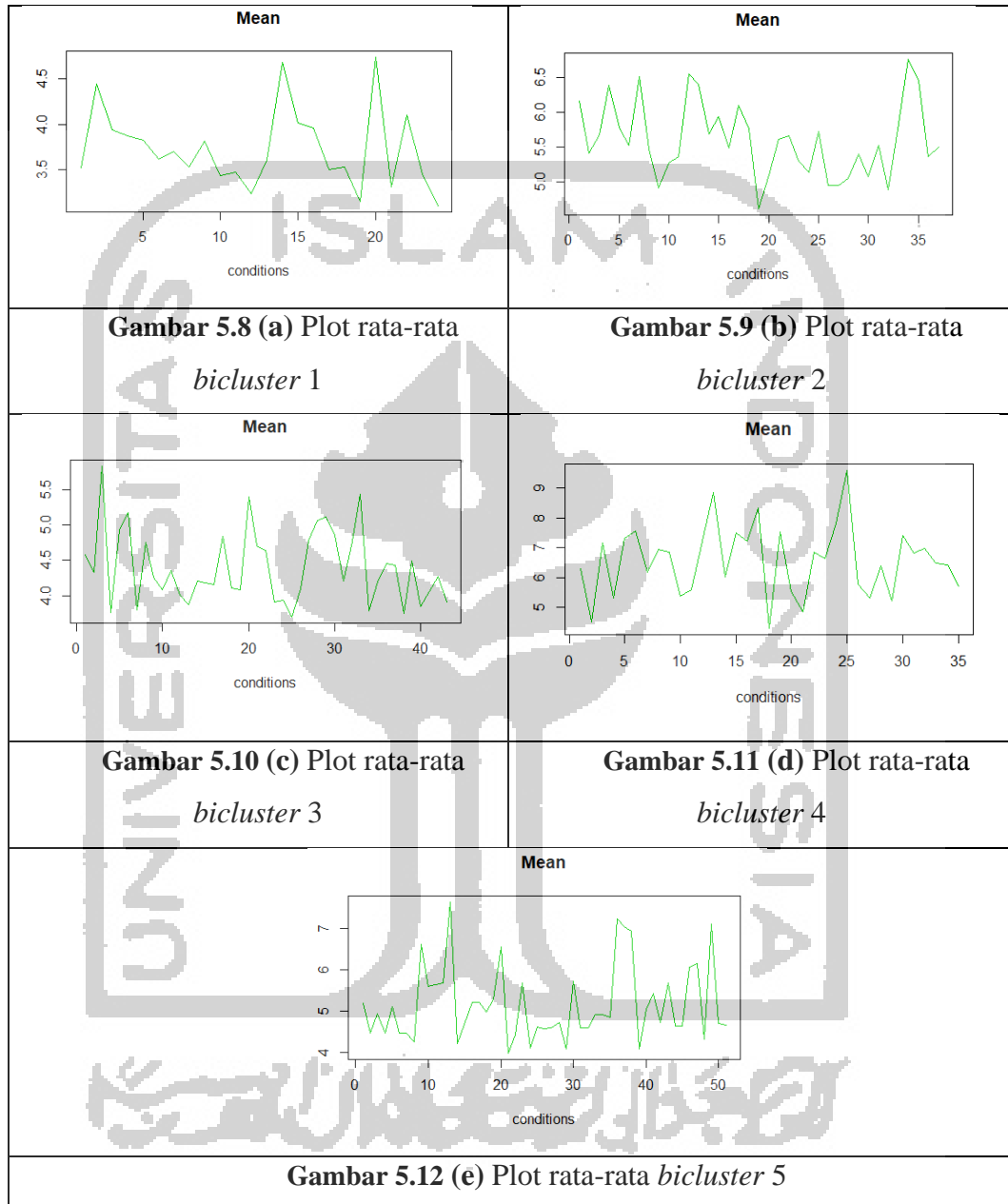
Pada Berdasarkan **Gambar 5.6** diketahui terdapat gen yang memiliki nilai ekspresi yang tinggi. Gen tersebut adalah gen X231375\_x\_at, X1553573\_at, X214041\_x\_at, X213350\_at. Hal tersebut di tunjukkan dengan warna merah sedikit gelap pada *heatmap*.



**Gambar 5.7** Heatmap Biclustering 5

Pada Berdasarkan **Gambar 5.3** diperoleh informasi bahwa sebagian besar organ memiliki ekspresi gen relatif sama. Sedangkan gene X222341\_at dan

X214149\_s\_at memiliki tingkat ekspresi gen yang tinggi. Hal tersebut di tunjukkan dengan warna sedikit gelap pada *heatmap*.



**Gambar 5.8 (a)** adalah plot nilai rata-rata ekspresi gen pada *bicluster1*. *Bicluster 1* memiliki rata-rata tertinggi pada gen X231500\_s\_at dan X223657\_at, sedangkan ekspresi gen yang memiliki rata-rata terendah yaitu X229589\_x\_at. pada *bicluster 1* telah diketahui nilai rata-rata pada setiap *bicluster* tertinggi X223657\_at. Maka gen X223657\_at hendaknya mendapatkan perhatian hal ini disebabkan nilai



ekspresi gen yang paling tinggi di *cluster* 1 dibandingkan gen yang lain, sedangkan untuk gen X229589\_x\_at merupakan gen dengan nilai ekspresi gen terendah sehingga prioritas penanganan pada gen tersebut cenderung dapat dikesampingkan dalam penanganan pada pasien

**Gambar 5.8 (b)** merupakan plot nilai rata-rata ekspresi gen pada *bicluster2*. Pada *bicluster2* terdapat gen yang memiliki nilai rata-rata ekspresi gen tertinggi yaitu gen X219138\_at sedangkan gen yang memiliki nilai ekspresi terendah pada gen X221986\_s\_at. Maka gen X219138\_at hendaknya mendapatkan perhatian hal ini disebabkan nilai ekspresi gen yang paling tinggi di *cluster* 2 dibandingkan gen yang lain, sedangkan untuk gen X221986\_s\_at merupakan gen dengan nilai ekspresi gen terendah sehingga prioritas penanganan pada gen tersebut cenderung dapat dikesampingkan dalam penanganan pada pasien.

**Gambar 5.8 (c)** adalah plot nilai rata-rata ekspresi gen pada *bicluster3*. Pada *bicluster3* terdapat gen yang memiliki nilai rata-rata ekspresi gen tertinggi yaitu gen X230790\_x\_at sedangkan gen yang memiliki nilai ekspresi terendah pada gen X240990\_at. Maka gen X230790\_x\_at hendaknya mendapatkan perhatian hal ini disebabkan nilai ekspresi gen yang paling tinggi di *cluster* 3 dibandingkan gen yang lain, sedangkan untuk gen X240990\_at merupakan gen dengan nilai ekspresi gen terendah sehingga prioritas penanganan pada gen tersebut cenderung dapat dikesampingkan dalam penanganan pada pasien

**Gambar 5.8 (d)** adalah plot nilai rata-rata ekspresi gen pada *bicluster4*. Pada *bicluster4* terdapat gen yang memiliki nilai rata-rata ekspresi gen tertinggi yaitu gen X213350\_at sedangkan gen yang memiliki nilai ekspresi terendah pada gen X228993\_at. Maka gen X213350\_at hendaknya mendapatkan perhatian hal ini disebabkan nilai ekspresi gen yang paling tinggi di *cluster* 4 dibandingkan gen yang lain, sedangkan untuk gen X228993\_at merupakan gen dengan nilai ekspresi gen terendah sehingga prioritas penanganan pada gen tersebut cenderung dapat dikesampingkan dalam penanganan pada pasien

**Gambar 5.8 (e)** adalah plot nilai rata-rata ekspresi gen pada *bicluster5*. Pada *bicluster5* terdapat gen yang memiliki nilai rata-rata ekspresi gen tertinggi yaitu gen X224321\_at sedangkan gen yang memiliki nilai ekspresi terendah pada gen

X206636\_at. Maka gen X224321\_at hendaknya mendapatkan perhatian hal ini disebabkan nilai ekspresi gen yang paling tinggi di *cluster* 5 dibandingkan gen yang lain, sedangkan untuk gen X206636\_at. merupakan gen dengan nilai ekspresi gen terendah sehingga prioritas penanganan pada gen tersebut cenderung dapat dikesampingkan dalam penanganan pada pasien

**Tabel 5.4** karakteristik gen tertinggi tiap *bicluster*

<i>Bicluster</i>	Gene	Rata-rata
<i>Bicluster1</i>	X226637_at	4.74368
<i>Bicluster2</i>	X219138_at	6.76696
<i>Bicluster3</i>	X230790_x_at	5.831358
<i>Bicluster4</i>	X213350_at	9.592093
<i>Bicluster5</i>	X224321_at	7.633602

Dari **Tabel 5.4** kolom pertama memuat informasi kelas dari *bicluster* pada data, kolom ke-dua memuat informasi mengenai nama dari gen pada data, kolom ke-tiga memuat informasi mengenai nilai rata-rata dari ekspresi gen, kolom ke-empat memuat informasi mengenai letak gen pada data yang dianalisis **Tabel 5.4** menunjukkan informasi bahwa nilai rata-rata ekspresi gen yang tertinggi ada pada *bicluster* 4, yaitu gen X213350\_at. Dikarenakan gen yang terbentuk pada *bicluster* 4 memiliki nilai ekspresi yang tinggi pada tiap gen, sehingga gen X213350\_at adalah gen yang perlu diperhatikan dalam menangani pasien. Kelompok gen yang terkelompokan bersamaan dengan sampel pada masing masing kluster terdapat pada **Lampiran 2** secara lebih lengkap.