



Gene Signature Analysis

Projeto Final

Natal-RN

Junho de 2015

Sumário

1	Como Compilar	3
1.1	Requisitos	3
1.2	Gerar executável	3
2	Como Executar	3
2.1	Instalação	3
2.2	Requisitos	3
2.3	Como Executar	3
3	Como Usar	4

1 Como Compilar

1.1 Requisitos

Não é necessário o uso de uma IDE, porém é recomendado o uso da IDE *Eclipse Luna*. Deve ser usado o Java 8.

1.2 Gerar executável

Caso não esteja usando uma IDE, siga o tutorial:

<http://pt.wikihow.com/Criar-um-Arquivo-JAR>

Caso esteja usando o *Eclipse Luna* clique em *File* no menu superior, depois em *Export* e finalmente em *Runnable JAR File*.

2 Como Executar

2.1 Instalação

Não há necessidade de instalação do software/ferramenta.

2.2 Requisitos

Desenvolvido para ser rodado em qualquer sistema operacional.

2.3 Como Executar

Para executar basta clicar duas vezes no arquivo executável jar: *genesigno.jar* já criado pelos desenvolvedores do sistema.

Caso queira rodar pelo terminal, entre no diretório do executável, basta inserir o comando: \$

```
java -jar genesigno.jar
```

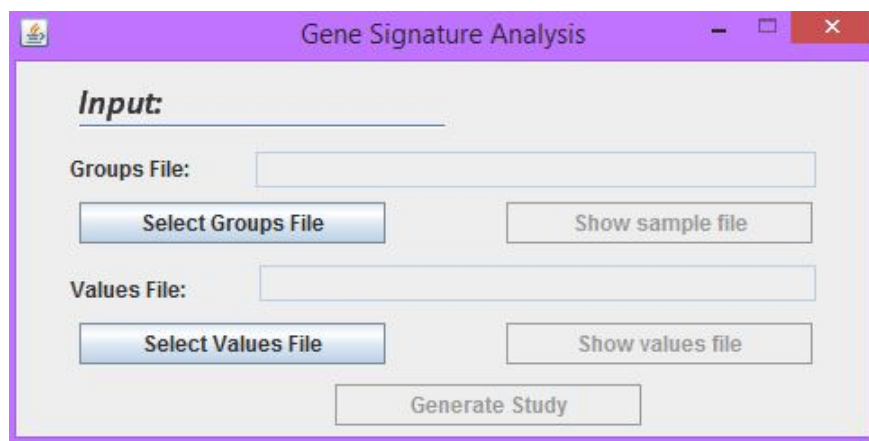
3 Como Usar

O programa possui uma tela inicial: interface principal:

The screenshot shows the main window of the 'Gene Signature Analysis' software. The window has a purple title bar with standard Windows window controls. The interface is divided into three main sections: 'Input:', 'Gene Signature:', and 'Clustering:'.
Input: This section contains two file selection areas. The first is for 'Groups File', with a text box and buttons for 'Select Groups File' and 'Show sample file'. The second is for 'Values File', with a text box and buttons for 'Select Values File' and 'Show values file'. Below these is a 'Generate Study' button and a list of output types: 'Groups', 'Samples', and 'Genes'.
Gene Signature: This section is for statistical analysis. It includes a 'Choose Statistical test:' label with four radio button options: 'Mann-Whitney U' (selected), 'Student's t', 'Kolmogorov-Smirnov', and 'One-Way ANOVA'. There is a 'p-value threshold' label with a text box containing '0.05' and a 'Gene Signature' button.
Clustering: This section is for hierarchical clustering. It includes a 'Choose Metric:' label with four radio button options: 'Euclidian' (selected), 'Manhattan', 'Square Euclidian', and 'Maximum'. Below this is a 'Choose Linkage Criteria:' label with three radio button options: 'Complete Linkage' (selected), 'UPGMA', and 'Single Linkage'. At the bottom of this section are two buttons: 'Generate Matrix' and 'Cluster'.

Ao abrir o programa é necessário que primeiramente seja inserido os dois arquivos de texto com os dados de expressão normalizados.

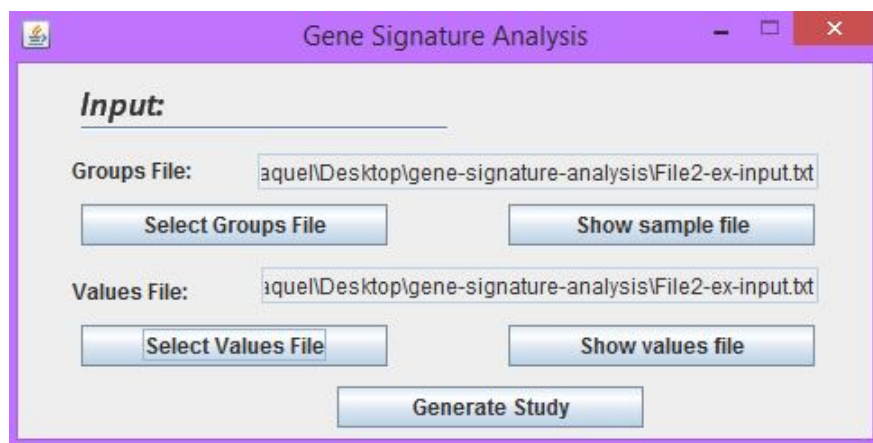
Clique em *Select Sample File* para selecionar o arquivo que informa os grupos correspondentes as amostras; ao selecionar o arquivo o diretório no qual o arquivo se encontra pode ser visualizado no campo ao lado de *Sample File*. Para visualizar e/ou verificar o arquivo selecionado, basta clicar em *Show Sample file*.



The screenshot shows the 'Gene Signature Analysis' window. Under the 'Input:' section, there are two rows of controls. The first row is for 'Groups File', featuring a text input field, a 'Select Groups File' button, and a 'Show sample file' button. The second row is for 'Values File', featuring a text input field, a 'Select Values File' button, and a 'Show values file' button. At the bottom center of the input section is a 'Generate Study' button.

Clique em *Select Values File* para selecionar o arquivo que informa a expressão dos genes de cada amostra; ao selecionar o arquivo o diretório pode ser visualizado no campo ao lado de *Values File*.

Para visualizar e/ou verificar o arquivo selecionado, basta clicar em *Show Values file*.



This screenshot shows the same 'Gene Signature Analysis' window, but now the text input fields contain file paths. The 'Groups File' field and the 'Values File' field both contain the path 'aque\l\Desktop\gene-signature-analysis\File2-ex-input.txt'. The buttons 'Select Groups File', 'Show sample file', 'Select Values File', 'Show values file', and 'Generate Study' remain visible in their respective positions.

Após os dois arquivos serem selecionado o botão *Generate Study* é habilitado e após clicado um *estudo* é gerado. Quando o *estudo* é gerado, aparecerá novas opções para que possa ser gerada uma assinatura genica.

Gene Signature Analysis

Input:

Groups File: \aque\lDesktop\gene-signature-analysis\File2-ex-input.txt
 Select Groups File Show sample file

Values File: \aque\lDesktop\gene-signature-analysis\File1-ex-input.txt
 Select Values File Show values file

Generate Study

2	Groups
10	Samples
1000	Genes

Gene Signature:

Choose Statistical test:

☒ Mann-Whitney U ☐ Kolmogorov-Smirnov
☐ Student's t ☐ One-Way ANOVA

p-value threshold 0.05

Gene Signature

De acordo com a quantidade de grupos existentes será possível escolher o teste estatístico. Caso possua mais de dois grupos o único teste disponível é o *Test One*, caso trate-se dois grupos é possível escolher entre: *Student t*, *Mann-Whitney U* e o *Kolmogorov-Smirnov*.

É possível modificar o p-valor para o critério de corte alterando o valor na caixa de texto ao lado de *p-value threshold*. Para gerar a assinatura genica sem os genes não informativos, basta apertar no botão *Gene Signature*.

Quando a assinatura é gerada outra tela aparece com essa assinatura genica e uma opção de salvar através do botão *Save Gene Signature*.

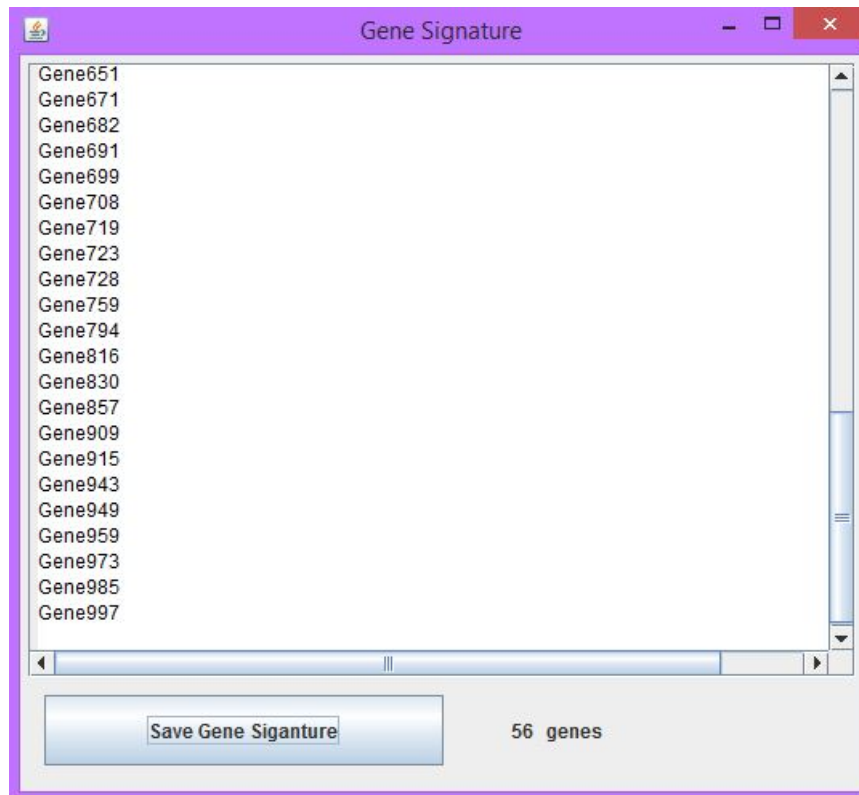


Figura 1: Simulação de uma assinatura gênica

Para continuar o processo você pode fechar a tela de assinatura genica, ou não, e será possível gerar uma matriz de distâncias clicando no botão *Generate Matrix*.

Ao gerar a matrix de distâncias uma tela abrirá mostrando a mesma e uma opção de salvar através do botão *Save Distance Matrix*.

	CT1	CT2	CT3	CT4	CT5	c6	c7
CT1	0,00	1110,46	1135,47	1186,99	912,77	1816,81	
CT2	1110,46	0,00	1152,69	1323,86	1037,17	1844,47	
CT3	1135,47	1152,69	0,00	1262,84	1023,47	2048,49	
CT4	1186,99	1323,86	1262,84	0,00	1034,22	1812,93	
CT5	912,77	1037,17	1023,47	1034,22	0,00	1715,30	
c6	1816,81	1844,47	2048,49	1812,93	1715,30	0,00	
c7	1977,59	2016,23	2083,03	1715,72	1766,07	1340,57	
c8	1960,99	1989,84	2213,14	2006,38	1924,57	1081,17	
c9	1904,83	1871,34	2066,04	1693,23	1708,35	1374,87	

Figura 2: Simulação da exibição da matrix de distâncias

Para fazer o agrupamento você pode modificar o critério de ligação ou permanecer com o critério padrão, *Complete Linkage*, e finalmente clicar no botão *Cluster*.

The screenshot shows the 'Gene Signature Analysis' window. It is divided into three main sections: 'Input:', 'Gene Signature:', and 'Clustering:'.
In the 'Input:' section, there are text boxes for 'Groups File' and 'Values File', both containing file paths. Below each text box are buttons for 'Select Groups File', 'Show sample file', 'Select Values File', and 'Show values file'. A 'Generate Study' button is also present. To the right of this button, the parameters '2 Groups', '10 Samples', and '1000 Genes' are listed.
The 'Gene Signature:' section contains a 'Choose Statistical test:' label with four radio button options: 'Mann-Whitney U' (selected), 'Student's t', 'Kolmogorov-Smirnov', and 'One-Way ANOVA'. Below these is a 'p-value threshold' label with a text box containing '0.05' and a 'Gene Signature' button.
The 'Clustering:' section has a 'Choose Metric:' label with four radio button options: 'Euclidian' (selected), 'Manhattan', 'Square Euclidian', and 'Maximum'. Below this is a 'Choose Linkage Criteria:' label with three radio button options: 'Complete Linkage' (selected), 'UPGMA', and 'Single Linkage'. At the bottom of this section are two buttons: 'Generate Matrix' and 'Cluster'.

Quando o botão for clicado, o agrupamento é realizado e uma nova tela será visualizada com o dendograma gerado e uma opção para salvar a árvore que representa o dendograma através do botão *Save tree*, uma opção de salvar a imagem do dendograma através do botão *Save Dendogram* e uma opção de validação através do botão *Validade*.

