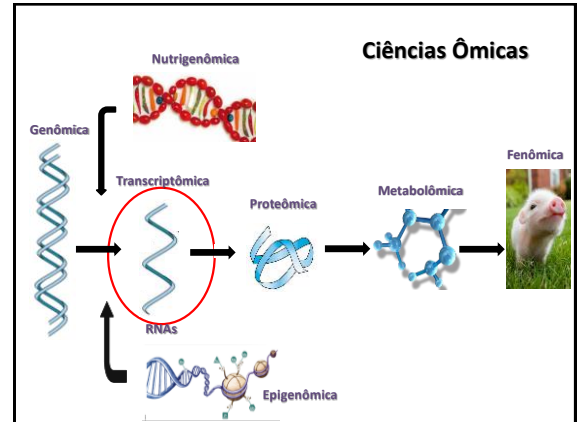


Aplicabilidade de Softwares em Análises Genômicas UFBA



Análises práticas de transcriptoma (RNA-Seq)

Dra. Larissa Fernanda Simielli Fonseca
Zootecnista
Pós-doutoranda FCAV/UNESP
PPG Genética e Melhoramento Animal



Transcriptoma

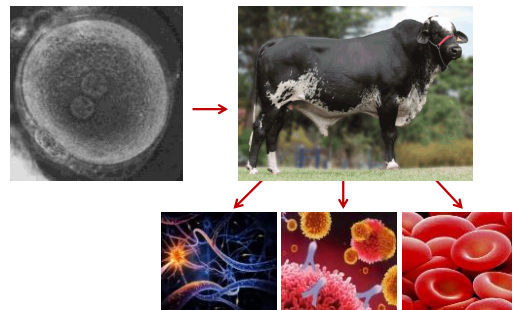
- A soma de todas as moléculas de RNA
 - produzidas em uma célula
 - sob um determinado conjunto de condições
- O seu estudo permite:
 - detecção de genes
 - mecanismos moleculares



podem ajudar na descoberta de biomarcadores para características de interesse econômico

Leininger, 14 ed
Cassar-Matell, 2008

Como diferentes tipos de células podem ser originadas de uma única célula fecundada?



Ranking	Pâncreas	%	Fígado	%
1	Procarboxipeptidase A1	7,6	Albumina	3,5
2	Tripsinogênio	5,5	Apolipoproteína A-I	2,8
3	Quimotripsinogênio	4,4	Apolipoproteína C-I	2,5
4	Tripsina	3,7	Apolipoproteína C-III	2,1
5	Elastase	2,4	ATPase	1,5
6	Protease E	1,9	Citocromo oxidase 3	1,1
7	Lipase	1,9	Citocromo oxidase 2	1,1
8	Procarboxipeptidase B	1,7	Alfa1-antitripsina	1,0
9	Amilase Pancreática	1,7	Citocromo oxidase 1	0,9
10	Lipase estimulada por sais biliares	1,4	Apolipoproteína E	0,9

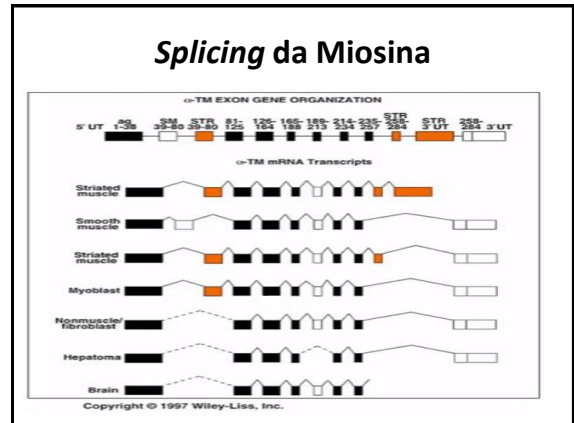
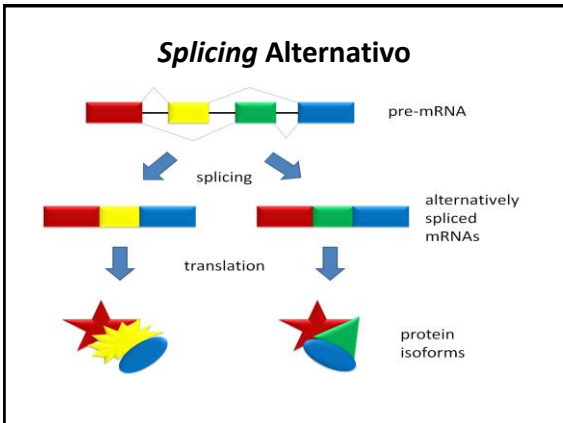
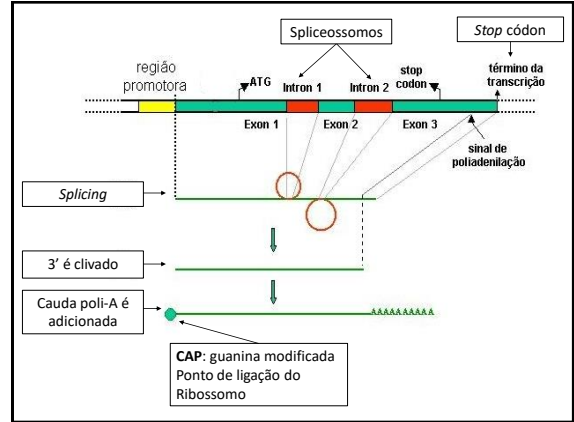
Complexidade dos Genomas Eucariotos

- Genomas grandes
 - Aproximadamente 22 mil genes em bovinos
- Divididos em cromossomos
- Diferentes tipos celulares
 - Regiões ativas diferentes
 - Resultam em transcritos diferentes

- Após a transcrição do RNAm, são necessárias algumas alterações

Pré RNA → RNA maduro

- *Splicing*
- Adição do CAP e Cauda Poli-A



RNA não codificador

- Molécula de RNA funcional, ou seja, não precisa ser traduzida em proteína para que a informação contida em sua sequência exerça sua função
- RNA não codificadores (ncRNA) são todos os RNAs que não são RNA mensageiros (mRNA), portanto, que não codificam proteínas

- **Tipos:**
 - tRNA
 - rRNA
 - snoRNA
 - snRNA
 - microRNA

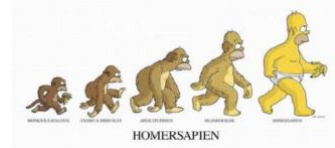
- Porção do transcriptoma que não é codificadora chega a ser até **4 vezes** maior que a porção codificadora
- Maior parte do genoma, que não codifica proteínas, parece estar envolvido na regulação da expressão gênica durante o **desenvolvimento** e a **diferenciação** dos organismos mais complexos

Sequenciamento de DNA

- É o processo de determinação da ordem precisa de nucleotídeos na molécula de DNA
- Inclui qualquer método ou tecnologia que é usada para determinar a ordem das quatro bases nitrogenadas: Adenina, Guanina, Citosina e Timina
- Com o advento de métodos rápidos de sequenciamento de DNA, grandes descobertas e pesquisas médicas e biológicas tem sido aceleradas na última década

Para que sequenciar?

O que diferencia as formas de vida é a seqüência, organização e expressão do material genético



Objetivos para sequenciamento genômico

Objetivos para sequenciamento genômico		
		Exemplo
Sequenciamento <i>de novo</i>	Sequenciamento genômico	Sequenciamento de >1000 genomas de influenza
	DNA de org. extinto	<i>Homo neanderthalensis</i>
	Metagenômica	Intestino humano
Resequenciamento	Genomas completos	Indivíduos humanos
	Regiões genômicas	Deteção de rearranjos ou regiões associados à doenças
	Mutações somáticas	Em câncer
Transcriptoma	mRNA	Definir regulação da transcrição
	Serial Analysis of Gene Expression (SAGE)	
	RNAs não codificadores	Identificar e quantificar microRNAs
Epigênética	Padrão de Metilação	Avaliar padrão de metilação em câncer

Como Sequenciar?

- **Sequenciamento de Primeira Geração (CLONAGEM)**
 - Degradação química – Maxam & Gilbert
 - Interrupção da cadeia (ddNTPs) – Sanger
- **Sequenciamento de Segunda Geração (AMPLIFICAÇÃO CLONAL)**
 - HiSeq, MiSeq, HiScan SQ - Illumina
 - 454 –Roche
 - Solid – Applied Biosystem
 - Ion Torrent – ABI – Life Technologies
- **Sequenciamento de Terceira Geração (SINGLE MOLECULE)**
 - Nanopore – Gridlon/Minilon
 - Heliscope - Helicos Biosciences
 - PacBio RS – Pacific Biosciences

2ª Geração – Amplificação Clonal




Illumina



Ion Torrent

Tecnologias

Plataforma	Preparo das Amostras	Comprimento das Leituras	Rendimento por Corrida	Método de Sequenciamento	Precisão
Illumina Solexa	PCR em Fase Sólida	300 pb	600 Gb	Sequenciamento por síntese	>98,5%
Ion Torrent	PCR em Emulsão	400 pb	10 Gb	Sequenciamento por síntese	98%



PLATFORM	ILLUMINA HISEQ 2500	ION PROTON
Library preparation	Bridge amplification	emPCR
Sequencing chemistry	Reversible dye terminators	pH change
Read length	From 2x125 bp to 2x300 bp	Up to 200 bp
Run time	7 hrs-6 days	From 2 to 4 hrs
Throughput/run	500-1000Gb (1Tb)	10Gb (P1), 100Gb (P1I)
Equipment Cost	750,000 \$	250,000 \$
Reagents Cost/run	5,500 \$	1,000 \$
GOOD!	High throughput/low cost per base/ease of use	Quick, easy to use and cheap
BAD!	Short sequences Strand-specific errors substitutions towards the end of the read, base substitution errors (systematic error GGT XGGG)	Errors in homopolymers Higher bias than Illumina <small>Wise et al. Genome Biology 2013, 14:R53 http://genomebiology.com/2013/14/R53</small>

Qual a principal vantagem do sequenciamento de segunda geração?

TEMPO!!!

O sequenciamento do genoma humano demorou 13 anos para ser concluído
Com essa nova tecnologia, sequenciar o genoma humano demora em média 1 semana!

Etapas

- Preparo das Bibliotecas de cDNA
- Amplificação clonal
 - PCR em emulsão
 - PCR em fase sólida
- Sequenciamento

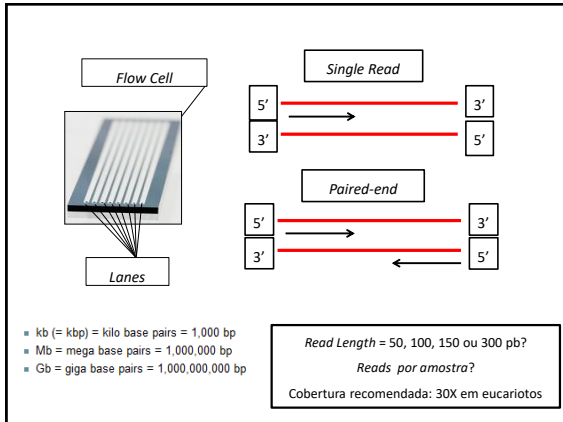
Metodologias – Sequenciamento de 2ª Geração

Illumina

- HiScan SQ: Híbrido (Sequenciamento e Genotipagem)
- HiSeq: sequenciamento em alta escala
- MiSeq: sistema de pequena capacidade




	MiSeq	NextSeq	HiSeq 2500	HiSeq X Ten
Output	15 Gb	120 GB	1000 GB	1800 GB
Number of Reads	25 Million	400 Million	4 Billion	6 Billion
Read Length	2x300 bp	2x150 bp	2x125 bp (2x250 update mid-2014)	2x150 bp
Cost	\$99K	\$250K	\$740K	\$10M

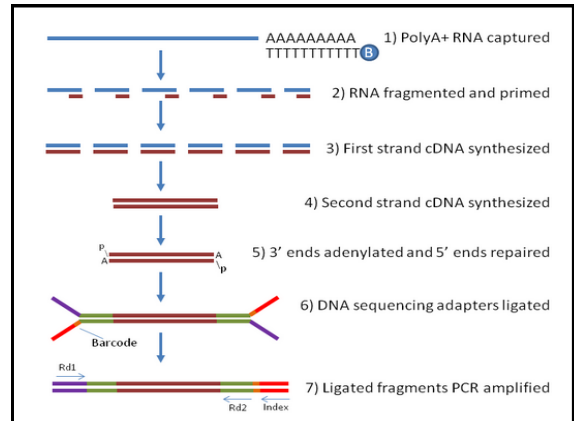


Metodologia Illumina

• Passos:

- 1) Isolamento do RNA mensageiro, montagem das bibliotecas de cDNA e ligação dos adaptadores
- 2) Clusterização das bibliotecas de cDNA na *flow cell*
- 3) Sequenciamento das amostras

1) Isolamento do RNA mensageiro, montagem das bibliotecas de cDNA e ligação dos adaptadores



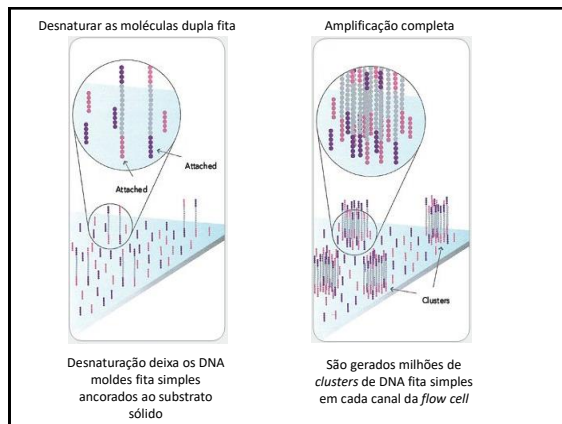
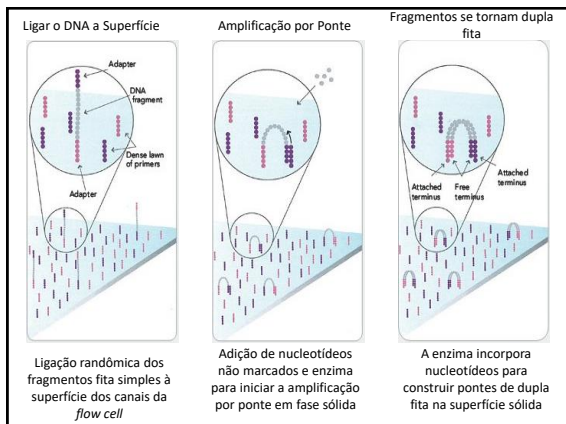
Validação

- Bioanalyzer: bibliotecas com 260 pb

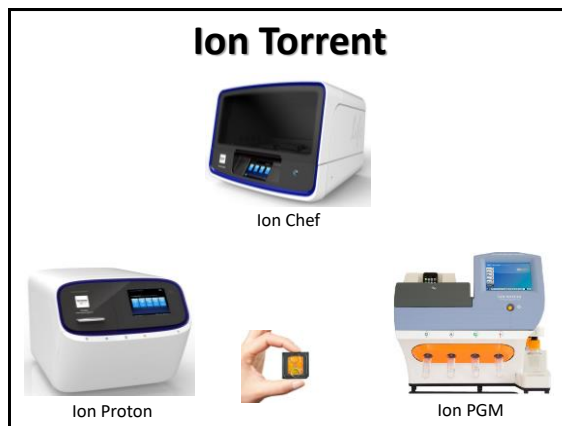
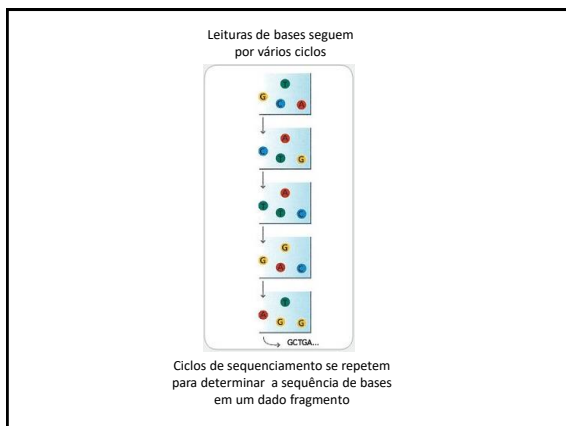
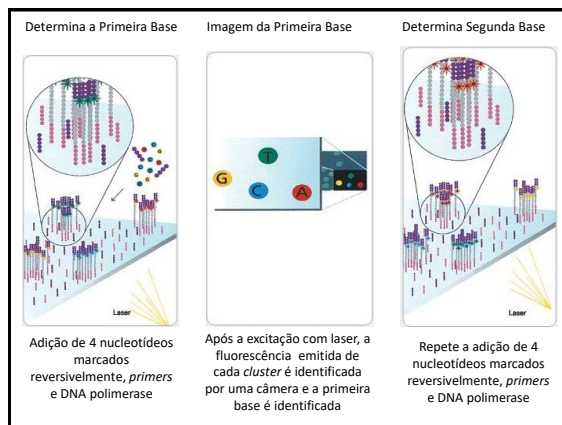
Normalização

- Qubit: quantificação
- 1ª Diluição: concentração de 10 nM
- Qubit: quantificação
- 2ª Diluição: concentração de 1,5 nM

2) "Clusterização" das bibliotecas de cDNA na *flow cell*



3) Sequenciamento das amostras

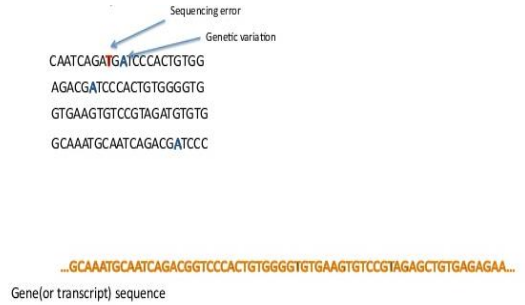


Alinhamento com genoma de Referência X De Novo Assembly

Combinação dos 2 métodos:

- Caso não possua um genoma de referência, pode-se montar o genoma de um organismo usando como referência um organismo filogeneticamente próximo
- Recomendado alinhar com as 2 metodologias, mesmo que tenha o genoma de referência

Alinhamento com genoma de Referência



Alinhamento com genoma de Referência

AGACGATCCCACTGTGGGGTG
GTGAAGTGTCCGTAGATGTGTG
GCAAATGCAATCAGACGATCCC

CAATCAGATGATCCCACTGTGG
...GCAAATGCAATCAGACGGTCCCACCTGTGGGGTGTGAAGTGTCCGTAGAGCTGTGAGAGAA...

Alinhamento com genoma de Referência

GTGAAGTGTCCGTAGATGTGTG
GCAAATGCAATCAGACGATCCC

AGACGATCCCACTGTGGGGTG
CAATCAGATGATCCCACTGTGG
...GCAAATGCAATCAGACGGTCCCACCTGTGGGGTGTGAAGTGTCCGTAGAGCTGTGAGAGAA...

Alinhamento com genoma de Referência

GCAAATGCAATCAGACGATCCC

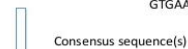
GTGAAGTGTCCGTAGAGCTGTG
AGACGATCCCACTGTGGGGTG
CAATCAGATGATCCCACTGTGG
...GCAAATGCAATCAGACGGTCCCACCTGTGGGGTGTGAAGTGTCCGTAGAGCTGTGAGAGAA...

De Novo Assembly

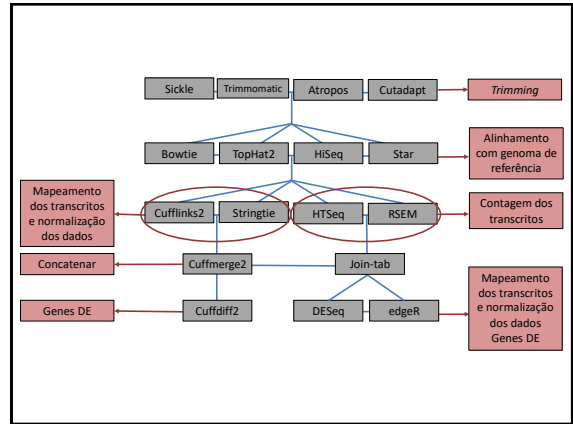
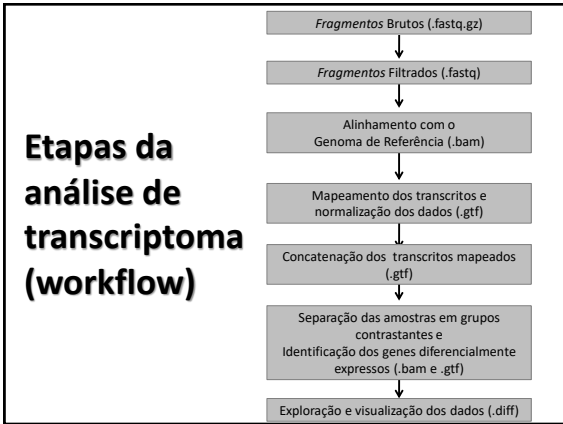
CAATCAGATGATCCCACTGTGG
AGACGATCCCACTGTGGGGTG
GTGAAGTGTCCGTAGATGTGTG
GCAAATGCAATCAGACGATCCC



CAATCAGATGATCCCACTGTGG
AGACGATCCCACTGTGGGGTG
GCAAATGCAATCAGACGATCCC
"singleton"
GTGAAGTGTCCGTAGATGTGTG



GCAAATGCAATCAGACGGTCCCACCTGTGGGGTG
GCAAATGCAATCAGATGATCCCACTGTGGGGTG



	Cuffdiff2	DESeq	edgeR
Distribuição	Normal	Binomial Negativa	Binomial Negativa
Teste estatístico	Teste-T	Teste Exato de Fisher	Teste Exato de Fisher
Estimação dos Transcritos	FPKM	Contagem das reads	Contagem das reads
Normalização dos Dados	Log2(FPKM)	Log2 (Expressão Reativa)	Média Aparada dos Valores de M
Dispersão	-	Método dos Momentos	Máxima Verossimilhança Condicional

Rodando as análises

- No seu próprio servidor
 - Linux
 - Instalar programas
 - Memória mínima necessária: 64 Gb RAM

Rodando as análises

- plataforma web de código fonte aberto
 - Galaxy Embrapa (<https://www.lmb.cnptia.embrapa.br/galaxy>)
 - Galaxy Europe - 3drnaseq app (https://3drnaseq.hutton.ac.uk/app_direct/3DRNaseq/#how-to-get-help)
 - CyVerse (<https://de.cyverse.org/de/>)

Rodando as análises

- Programas pagos
 - QIAGEN CLC Genomics Workbench (<https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-clc-genomics-workbench/>)

Na prática...

Protocolo Tuxedo

Plataforma CyVerse

- Criar *login* para acesso
- Fazer *upload* das sequências (.fastq.gz)
- Descompactar as sequências (fastq)
- Verificar qualidade das sequências
- Realizar limpeza (*trimming*)
- Verificar qualidade das sequências “trimadas”
- Alinhar com genoma de referência (.bam)
- Montar o genoma/transcriptoma de referência (.gtf)
- Encontrar genes DE (.bam e .gtf)

Alinhamento com Genoma de Referência

Discovery Environment

Analysis Name: TSPAC_HL_andr6G

Analysis Source: TSPAC_HL_andr6G

Select output folder

Launch

Instructions: Please upload to two FASTQ files for each set of paired-end reads. Left and right reads from separate read lanes. Click down to input right read file. NOTE: For multiple files, the left and right files must be in the same order.

Reference Genome (Optional)

Select a reference genome from the list or select your own reference genome file. Note one of these two options MUST be selected.

Select a reference genome from the list:

BSLIP002.1

Reference Annotation

Analysis Options

- k-mer size (k-mer): 31
- k-mer length: 31
- Maximum number of mismatches that can appear in the anchor region of spliced alignment: 3
- The minimum read length: 75
- The maximum read length: 3000
- Minimum number of segments to be aligned: 2
- Minimum number of segments that can be found during split-segment (left-to-right) search: 2
- Minimum number of segments that can be found during split-segment (right-to-left) search: 2
- Number of repetitive aligned in each segment alignment for each repeat independently: 3
- Minimum length of read segments: 25

Launch button

- **Cufflinks:**
 - mapeamento dos transcritos
 - Cálculo de FPKM
- **Cuffmerge**
 - Concatena todos os arquivos em um único transcriptoma que será usado como referência

Cuffdiff – genes DE

Name	Last Modified	Size
hsava_cuffdiff	2015-04-17 11:40:05	247 bytes
hsava_cuffdiff_info	2015-04-17 11:40:27	2,34 KB
cdfs_count_tracking	2015-04-17 11:41:12	12 bytes
cdfs_diff	2015-04-17 11:39:57	115 bytes
cdfs_diff_tracking	2015-04-17 11:41:02	96 bytes
cdfs_read_group_tracking	2015-04-17 11:41:28	110 bytes
cdfs_read_diff	2015-04-17 11:41:05	124 bytes
cuffData.db	2015-04-17 11:40:17	107,87 KB
genes_exp_diff	2015-04-17 11:41:07	1,38 KB
genes_count_tracking	2015-04-17 11:41:04	1,61 KB
genes_fpkms_tracking	2015-04-17 11:41:08	3,56 KB
genes_read_group_tracking	2015-04-17 11:40:18	97,18 KB
informs_exp_diff	2015-04-17 11:39:49	11,39 KB
informs_count_tracking	2015-04-17 11:40:08	4,61 KB
informs_fpkms_tracking	2015-04-17 11:40:05	11,3 KB
informs_read_group_tracking	2015-04-17 11:40:17	208,21 KB
promoters_diff	2015-04-17 11:39:45	2,12 KB
read_group_info	2015-04-17 11:40:47	1,11 KB
read_info	2015-04-17 11:38:49	903 bytes
splice_diff	2015-04-17 11:41:40	4,87 KB
tsv_group_exp_tracking	2015-04-17 11:40:23	5,68 KB
tsv_group_fpkms_tracking	2015-04-17 11:40:05	2,87 KB
tsv_group_read_group_tracking	2015-04-17 11:40:28	5,51 KB
tsv_read_info	2015-04-17 11:40:25	96,47 KB
tsv_read_info	2015-04-17 11:38:47	1,01 KB

Displaying 0 - 25 of 25

Exploração dos Dados: cummeRbund (ambiente R)

Linhas de comando

- Acesso a um computador com no mínimo 64 Gb de memória RAM
- Instalar todos os programas que serão utilizados e também as extensões e dependências
- Baixar e indexar o genoma de referência (.fa e .gtf)
- Sequências (.fastq.gz)
- Descompactar as sequências (fastq)
- Verificar qualidade das sequências
- Realizar limpeza (*trimming*)
- Verificar qualidade das sequências “trimadas”
- Alinhar com genoma de referência (.bam)
- Montar o genoma/transcriptoma de referência (.gtf)
- Encontrar genes DE (.bam e .gtf)

Scripts



- Prof. Daniel Guariz Pinheiro
Professor Assistente Doutor
Depto Tecnologia – FCAV/UNESP
- Informações e contato*
<https://www.fcav.unesp.br/#!/departamentos/tecnologia/docentes/daniel-quariz-pinheiro/main/>
- Disciplina ministrada para PG (Bioinformática Aplicada II: Análise de Transcriptomas)*
<https://www.fcav.unesp.br/#!/departamentos/tecnologia/docentes/daniel-quariz-pinheiro/teaching/graduate/>
- Scripts*
<https://github.com/dapinheiro/bioinfoutilities>

Scripts

- Limpeza dos dados**
 - Preprocess4**
 - Atropos** (sequências de baixa qualidade e adaptadores)
 - Prinseq** (filtrar, reformatar e trimar sequências)
- Alinhamento**
 - rnaseq.sh**
 - 4 opções: 2 alinhadores (**TopHat** ou **Star**) e 2 mapeadores (**stringtie** ou **cufflinks**)
 - Cuffmerge**: fusão dos arquivos gtf para referência
 - Cuffcompare**: comparar transcritos quando não há transcriptoma de referência
 - Cuffquant**: calcula os perfis de expressão de gene e transcrição e salva esses perfis em arquivos que você pode analisar posteriormente com **Cuffdiff** ou **Cuffnorm** (economizar RAM)
 - Cuffnorm**: "Nível extra de normalização" além do FPKM necessário em algumas situações
 - Cuffdiff**: comparar a regulação para cima ou para baixo entre duas ou mais condições

Resultados

Relatório fastqc

- fastqc SAMPLEA1_R1.fastq

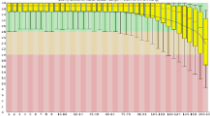
04/02/2021 SAMPLEA1_R1_FastQC Report Thu 4 Feb 2021
SAMPLEA1_R1.fastq

FastQC Report Summary

- ✓ Basic Statistics
- ✓ Per base sequence quality
- ✓ Per sequence quality scores
- ✓ Per base sequence content
- ✗ Per sequence GC content
- ✓ Per base N content
- ✓ Sequence Length Distribution
- ⓘ Sequence Duplication Levels
- ✓ Overrepresented sequences
- ✓ Adapter Content

Measure	Value
Filename	SAMPLEA1_R1.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.3
Total Sequences	29801
Sequences flagged as poor quality	0
Sequence length	151
Ntc	52

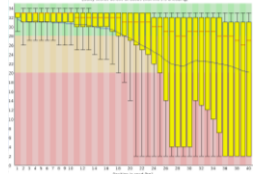
Per base sequence quality



Trimming


```

@HISEQ:68_ACAGTG_L007
ATCCTCTGCAGCCTCCAGCTCACTCTCAATGATGACCACTTAGCGGCCACTCTTCACTCTG
+
FB BB FFFFF FFFF FFFFF FFFF FFFF FFFF FFFF FFFF FFFF FFFF FFFF FFFF
    
```



→

Pré - processamento



←

**Pós - processamento
Trimming**

Alinhamentos

Left reads:

Input: 37254261
 Mapped: 33866989 (90.9% of input)
 of these: 3617188 (10.7%) have multiple alignments (16714 have >20)

Right reads:

Input: 37254261
 Mapped: 32620511 (87.6% of input)
 of these: 3314291 (10.2%) have multiple alignments (17119 have >20)
 89.2% overall read alignment rate.

Aligned pairs: 31432574

of these: 2488072 (7.9%) have multiple alignments
 and: 282210 (0.9%) are discordant alignments
 83.6% concordant pair alignment rate.

• Cufflinks:

- mapeamento dos transcritos
- Cálculo de FPKM
- Arquivo .gtf

• Cuffmerge

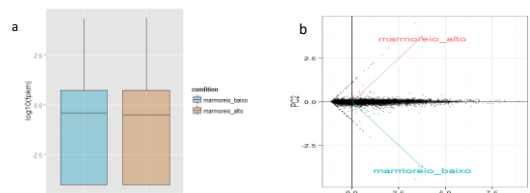
- Concatena todos os arquivos em um único transcrito que será usado como referência
- Grande arquivo .gtf formado a partir das saídas do Cufflinks

Genes DE

gene_id	gene	locus	sample_1	sample_2	status	value_1	value_2	log2(fold_change)	test_stat	p_value	q_value
ENSBTAG00000008157	IFT27	1434805043-58852086	male1_dura	maciez_macia_OK	OK	300.014	108.315	-14.698	-107.0418	0.0001	0.013573
ENSBTAG00000010279		6.8755283-25.23091496	male1_dura	maciez_macia_OK	OK	163.788	543.206	-139.226	-164.271	5.00E-05	0.00784688
ENSBTAG00000014836	ZCSCAN2	23104117-5.300388388	male1_dura	maciez_macia_OK	OK	0.689172	0.130967	-239.566	-122.734	5.00E-05	0.00784688
ENSBTAG00000017022	CLCC12A	120401253-217785972	male1_dura	maciez_macia_OK	OK	205.909	0.31943	-268.844	-11.035	5.00E-05	0.00784688
ENSBTAG00000017867	CLCC4G	17789026-5.71906699	male1_dura	maciez_macia_OK	OK	213.916	0.152397	-381.114	-136.7978	0.0002	0.02511
ENSBTAG00000020701	HMOX1	73987841-5.71906699	male1_dura	maciez_macia_OK	OK	115.512	197.874	0.795339	0.882776	0.0002	0.02511
ENSBTAG00000021428	SYP	92320415-3.204091409	male1_dura	maciez_macia_OK	OK	0.814972	0.169398	-226.804	-1.199	5.00E-05	0.00784688
ENSBTAG00000023107	CLDN19	130080449-1.30792451	male1_dura	maciez_macia_OK	OK	112.997	0.133369	-308.279	-137.899	5.00E-05	0.00784688
ENSBTAG00000023995	IQCG	70818105-5.94701433	male1_dura	maciez_macia_OK	OK	0.975199	0.0683124	-383.548	-157.571	5.00E-05	0.00784687
ENSBTAG00000024424		94302012-282050145	male1_dura	maciez_macia_OK	OK	215.314	0.19875	-343.882	-155.219	0.00035	0.0428707
ENSBTAG00000026100	ext.BOLA-DOB	25883052-1209994009	male1_dura	maciez_macia_OK	OK	122.213	682.692	-0.840093	-106.142	5.00E-05	0.00784688
ENSBTAG00000030611	DAMGD1	10606954	male1_dura	maciez_macia_OK	OK	0.700546	0.0730559	-330.145	-144.417	5.00E-05	0.00784688

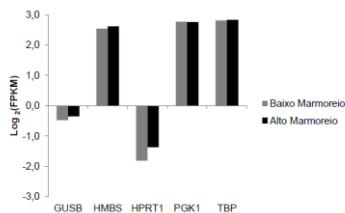
cummeRbund (ambiente R)

Exploração dos Dados



- (a) Boxplot do log10 do FPKM dos valores de expressão para os 2 grupos avaliados
 (b) Análise de componente principal (PCA) entre os 2 grupos avaliados

Comparação dos genes de referência



Na prática...

DeSeq e edgeR

até o alinhamento, seguimos da mesma forma...

DeSeq e edgeR

Contar reads: HTSeq (.txt)

No R – script edgeR

```
## Set up table as needed
rcount1 <- rcount[,c(6:25)]
names(rcount1) <-
c("HRFI0","HRFI1","HRFI2","HRFI3","HRFI4","HRFI5","HRFI6","HRFI7","HRFI8","HRFI9","LRFI0","LRFI1",
"LRFI2","LRFI3","LRFI4","LRFI5","LRFI6","LRFI7","LRFI8","LRFI9")

## Selecting only the animals classified for the traits
x <- rcount1[,c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20)]

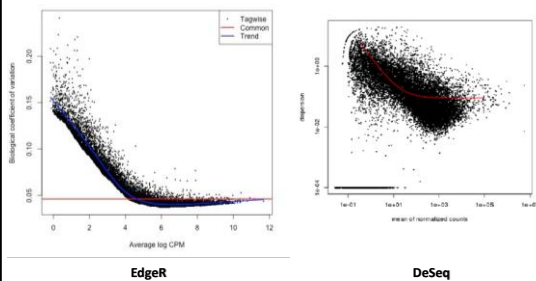
# Assign condition (first four are controls, second four contain the expansion)
trt <-
factor(c("HRFI","HRFI","HRFI","HRFI","HRFI","HRFI","HRFI","HRFI","HRFI","HRFI","LRFI","LRFI","LRFI",
"LRFI","LRFI","LRFI","LRFI","LRFI","LRFI"))

data.frame(sample=colnames(x), trt)
design <- model.matrix("0+trt")
design
```

DeSeq e edgeR

- Utiliza dados de contagem das *reads*
- Normalização realizada dentro dos programas
- Gráficos de dispersão
 - DeSeq:
 - Log2 (Expressão Reativa)
 - Método dos Momentos
 - edgeR:
 - Média Aparada dos Valores de M
 - Máxima Verossimilhança Condicional

Gráficos de Dispersão



edgeR output

Files:

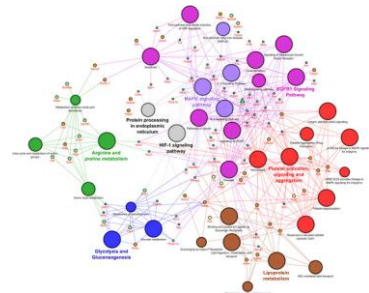
- **de-list-edger.tsv:** Tabela de resultados de testes estatísticos, incluindo estimativas de foldchange e valores p
 - logFC = log2 foldchange entre os grupos. Por exemplo, valor 2 significa que a expressão aumentou 4 vezes
 - logCPM = a média de log2 contagens por milhão
 - PValue = valor p
 - FDR = valor p ajustado
- **edger_report.pdf:** Um arquivo PDF contendo
 - ma-plot-edger.pdf: plotagem MA onde as características expressas de forma significativa são destacadas
 - dispersion-edger.pdf: Gráfico do coeficiente biológico de variação.
 - mds-plot-edger.pdf: Gráfico de escala multidimensional para visualizar semelhanças de amostra
 - p-value-plot-edger.pdf: Gráfico de distribuição do valor p bruto e ajustado
- **edger-log.txt:** Arquivo de log se nenhuma expressão significativamente diferente for encontrada

edgeR output

	gene_ID	logFC	logCPM	LR	Pvalue	FDR
6769	ENS000000133110	6.359164	10.952685	976.3333	2.503832e-214	4.605299e-210
8491	ENS000000143546	-5.544069	9.190249	762.0486	9.630025e-168	8.856253e-164
11057	ENS000000163220	-5.304425	10.255550	759.9131	2.805151e-167	1.719838e-163
8639	ENS000000144476	6.874676	5.656485	751.3280	2.063670e-165	9.489269e-162
5528	ENS000000124225	5.292294	8.728008	741.7265	2.525650e-163	9.290856e-160
10073	ENS000000156453	6.212131	5.829142	729.6710	1.056166e-160	3.237676e-157
3320	ENS000000106366	5.029834	13.904191	705.1793	2.235802e-155	5.874729e-152
2580	ENS000000101670	5.075895	7.026149	662.5557	4.154631e-146	9.552017e-143
13827	ENS000000174343	6.243975	5.039587	640.1431	3.110468e-141	6.356759e-138
12094	ENS000000166922	7.799969	4.681274	634.3902	5.546206e-140	1.020114e-136

Próximos passos

- Análise de enriquecimento



**Aplicabilidade de Softwares em
Análises Genômicas
UFBA**



Análises práticas de transcriptoma (RNA-Seq)

Obrigada!!!

Dra. Larissa Fernanda Simielli Fonseca
Zootecnista
lsimielli.fonseca@gmail.com

