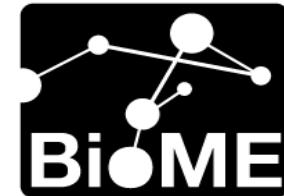


*In Company*

**CURSOS DE CURTA DURAÇÃO**

# **BIOINFORMÁTICA**

**BIOME - CENTRO MULTIUSUÁRIO DE BIOINFORMÁTICA - UFRN**



Curso teórico-prático

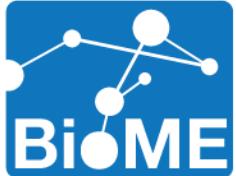
**INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO**

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020

## **Introdução**

**SANDRO J. DE SOUZA**

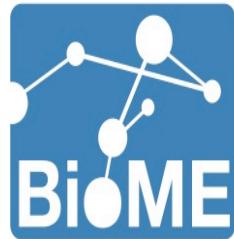


# BioME

INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020



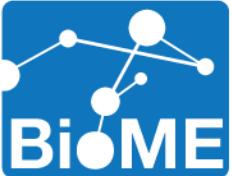
Bioinformatics  
Multidisciplinary  
Environment

Centro  
Multiusuário  
de Bioinformática



Fundado em 2016

<http://bioinfo.imd.ufrn.br>



# BioME

INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

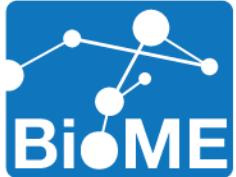
Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020

The slide features a white rectangular box containing logos and text. On the left is the BioME logo. To its right, separated by a vertical line, is the text "Bioinformatics Multidisciplinary Environment". Another vertical line separates this from the text "Centro Multiusuário de Bioinformática". Below these are three logos: "IME INSTITUTO METRÓPOLE DIGITAL", "PPG BIOINFO UFRN" (with a DNA sequence logo), and "UFRN UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE".

- ✓ "Minor" em Bioinformática
- ✓ MS/DR em Bioinformática
- ✓ Spinoff: DUNA Bioinformatics

<http://bioinfo.imd.ufrn.br>



# BioME

INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020



**Rodrigo Dalmolin**  
(Systems Biology)



**Jorge E. de Souza**  
(Genomics)



**César Renno-Costa**  
(Computational  
Neuroscience)



**Renan Moioli**  
(Computational  
Neuroscience)



**João Paulo Matos**  
(Genomics and Structural  
Bioinformatics)



**Sandro J. de Souza**  
(Genomics and Evolutionary  
Biology)



**Gustavo A. de Souza**  
(Proteomics)



**Beatriz Stransky**  
(Modelling of  
Biological Systems)



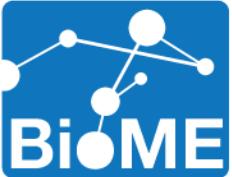
**Euzébio Barbosa**  
(Rational Drug  
Design)



**Daniel Sabino**  
(Machine Learning)



**Tetsu Sakamoto**  
(Phylogeny,  
Ontology)



# BioME

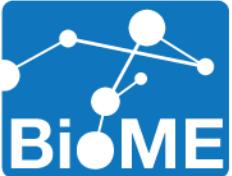
INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020

The collage includes:

- A top row showing three images: a large computer lab filled with students at desks, two researchers in lab coats working together on a laptop, and another computer lab with students and a presentation screen.
- A bottom row showing three images: a lecture hall full of students, the exterior of a modern building with "IMD INSTITUTO METRÓPOLE DIGITAL" and "BIOINFORMATICS" signs, and a computer lab with a teacher standing and a student presenting to a seated audience.
- A central white box containing the text "Cursos de Curta Duração" (Short Duration Courses).
- A blue call-to-action button with the text "Saiba Mais" (Learn More).

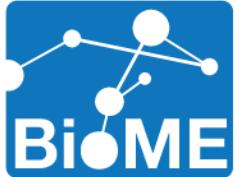


INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020

It is a great time to be  
a bioinformatician!



# INTRODUÇÃO À ANÁLISE DE DADOS DE SEQUENCIADORES DE SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020

educacao.estadao.com.br

SP 20°  
39° Acervo PME Jornal do Carro paladar link Radio Eldorado Radio Estadão in 8+ f Classificados ASSINE O ESTADÃO Buscar OUÇA AS RÁDIOS

## ESTADÃO

### Educação

ÚLTIMAS BLOGS COLUNAS

5 COMENTÁRIO(S)  
CLIQUE E DÊ A SUA OPINIÃO

ASSINE O ESTADÃO

# Profissões do Futuro

## Patrocínio: Senac

# Profissões do Futuro: bioinformática transforma dados em conhecimento

VIVIANE ZANDONADI - ESPECIAL PARA O ESTADO - O ESTADO DE S. PAULO  
30 Junho 2015 | 15h 09

Quem entende de biologia e computação ajuda a resolver os problemas das ciências da vida

Ao buscar na internet a palavra "bioinformacionista", descobrimos que ela não está no dicionário e tem sido usada em uma série de estudos e em exercícios de futurologia de carreiras, como este artigo, que tentam antecipar em que estaremos trabalhando daqui a dez ou vinte anos. O bioinformacionista seria o profissional do futuro da área de saúde. Alguém que a partir de informações biológicas apoia os cientistas da saúde na compreensão dos problemas e no desenvolvimento de modernos tratamentos e remédios.

ENCONTRE SEU IMÓVEL

MOVINGIMÓVEIS DIRETO AO PONTO

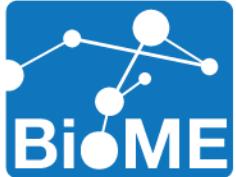
RECOMENDADAS

Blog Ser Mãe: A mídia diz que você precisa ser magra também no pós-parto

'Star Wars: O Despertar da Força' tem novo trailer e ingressos em pré-venda

Moradora inaugurou em SP o conceito das ruas que não abrem nos fins de semana

A Farra do Fies: especial



## INTRODUÇÃO À ANÁLISE DE DADOS DE SEQUENCIADORES DE SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020

**WIRED**

---

**SHARE**

---

SHARE 3738

---

TWEET

---

COMMENT

---

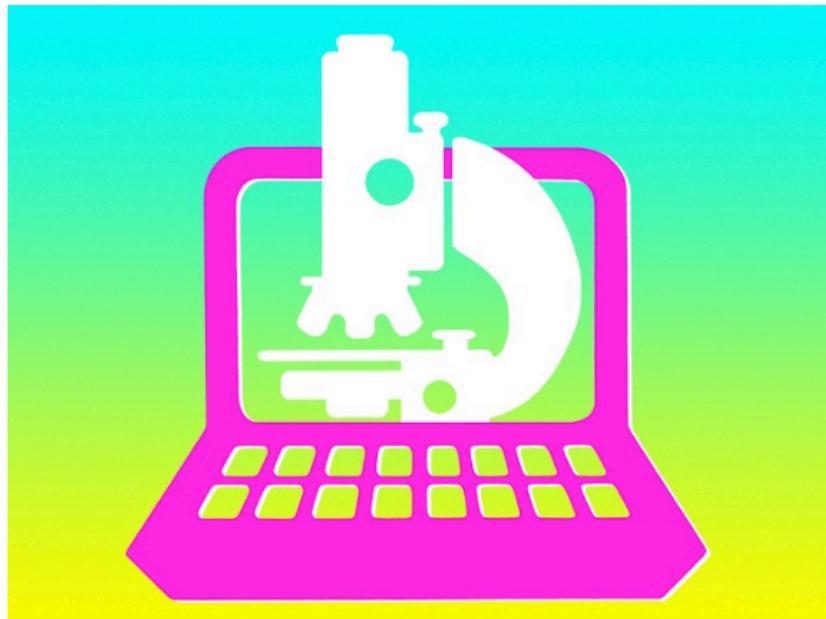
EMAIL

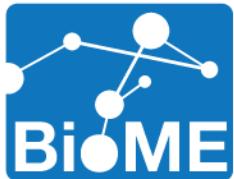
---

Want to Make It as a Biologist? Better Learn to Code

EMILY DREYFUSS SCIENCE 03.10.17 12:00 PM

# WANT TO MAKE IT AS A BIOLOGIST? BETTER LEARN TO CODE





Valor.com.br | ValorInveste | Valor RI | Cadastre-se | Login | Assine

# Valor ECONÔMICO

Princípios Editoriais

Home | Brasil | Política | Finanças | Empresas | Agronegócios | Internacional | Opinião | Legislação | Carreira | Cultura & Estilo | Valor Data

Suplementos & Revistas | Hoje no Valor | Blog do FMI | Especial Pequenas e Médias Empresas | Rumos da Economia 2018

## 27 de julho de 2018

ANTERIORES

**SUPLEMENTO - SAÚDE - TECNOLOGIA E INOVAÇÃO**

**Respostas rápidas**   
Uso de inteligência artificial na análise de grande volume de dados deve proporcionar diagnósticos e decisões médicas mais acertadas  
Equipamentos indicam diagnósticos e antecipam resultados



**Serviços Telemedicina dá agilidade ao atendimento**   
Cerca de 50 empresas especializadas apoiam seguradoras, operadoras, laboratórios e pacientes  
Teles fornecem serviços diferenciados de monitoramento



**Ferramentas promovem ganhos também para a gestão dos negócios**   
Busca por integração da tecnologia médica aos sistemas administrativos é crescente



**Realidade mista permite cirurgias complexas**   
Brasil dá os primeiros passos na tecnologia que une realidades virtual e aumentada

**Inovação Healthtechs atraem aporte de hospitais**   
Startups crescem com as iniciativas de aceleração e levam mercado a novo patamar  
Investidor busca ganho rápido de escala



**Automação Custo elevado inibe ritmo de avanço dos robôs cirúrgicos**   
País já realizou 17 mil cirurgias com auxílio do sistema da americana Intuitive Surgical  
Demanda por treinamento está em alta

**Plataformas de IoT geram soluções personalizadas**   
Fabricantes lançam equipamentos e aplicativos que dão conforto e praticidade a pacientes e médicos

**BNDES apoia fundo que investe em empresas da área de genômica**   
Brasil tem produção expressiva de trabalhos sobre pesquisa de sequenciamento genético

**Farmacêuticas buscam a inovação radical**   
Entraves burocráticos dificultam as pesquisas clínicas, a obtenção de insumos e o registro de patentes

**Anteriores**

**Assine o Valor**  
Edição Digital  
Digital + Impressa

**Filtrar por Editorias**

Índice Impresso  
Primeira página  
Brasil  
Política  
Internacional  
Opinião  
Especial  
Empresas  
Agronegócios  
Finanças  
EU & Fim de semana  
Legislação & Tributos  
Suplemento - Programas de fidelidade  
Suplemento - Saúde - Tecnologia e inovação  
Empresas Citadas

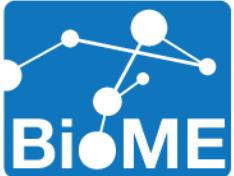
**SUA FAMÍLIA MERECE VIVER COM MUITO MAIS SEGURANÇA**



**SAIBA MAIS**  
**TOTALMENTE PRONTOS NA BARRA 2, 3 E 4 QUARTOS E COBERTURAS**

**Versão Digital**   
27-07-2018



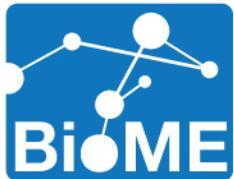


# Big Data

# INTRODUÇÃO À ANÁLISE DE DADOS DE SEQUENCIADORES DE SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020

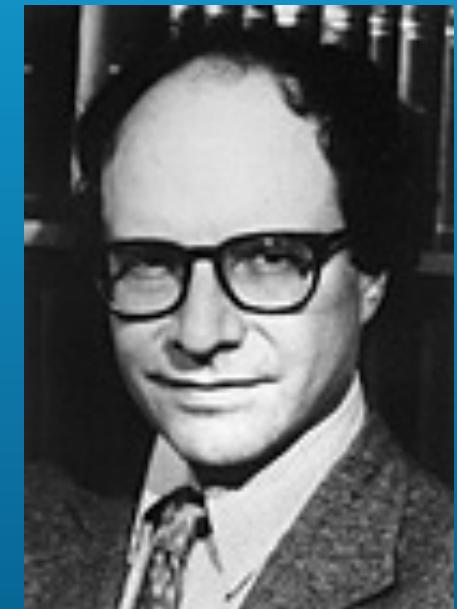
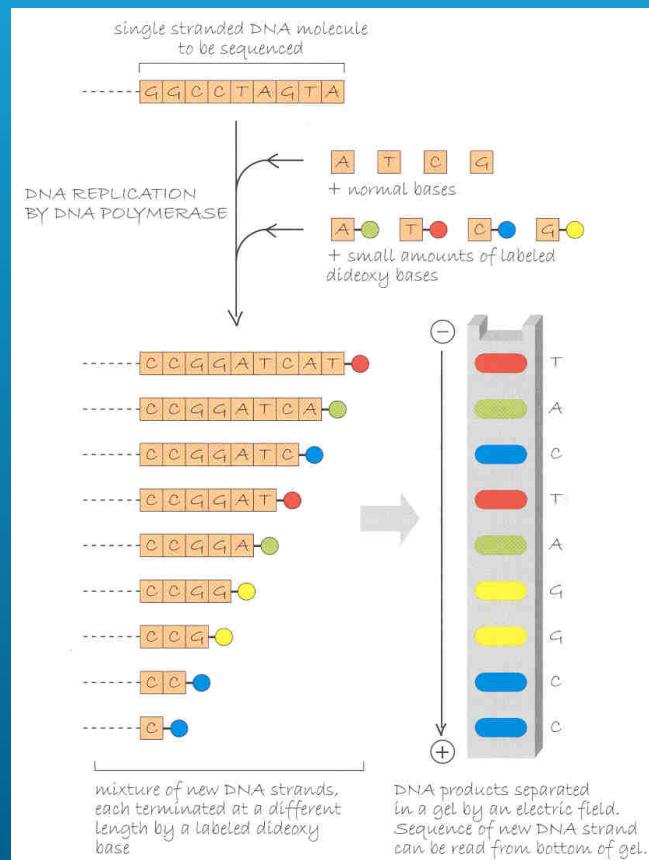
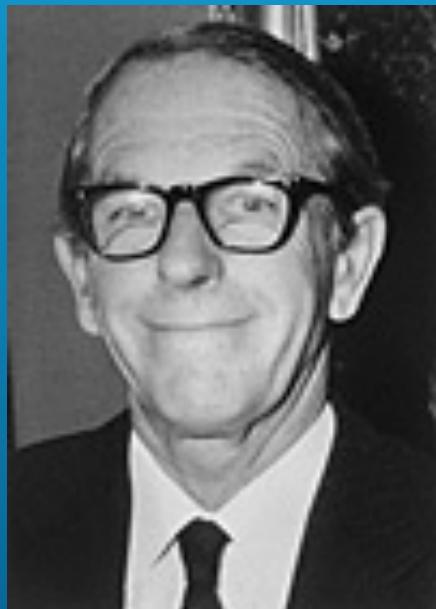


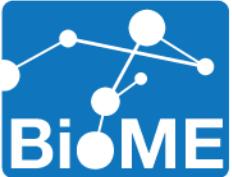
# DNA Sequencing

# INTRODUÇÃO À ANÁLISE DE DADOS DE SEQUENCIADORES DE SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020



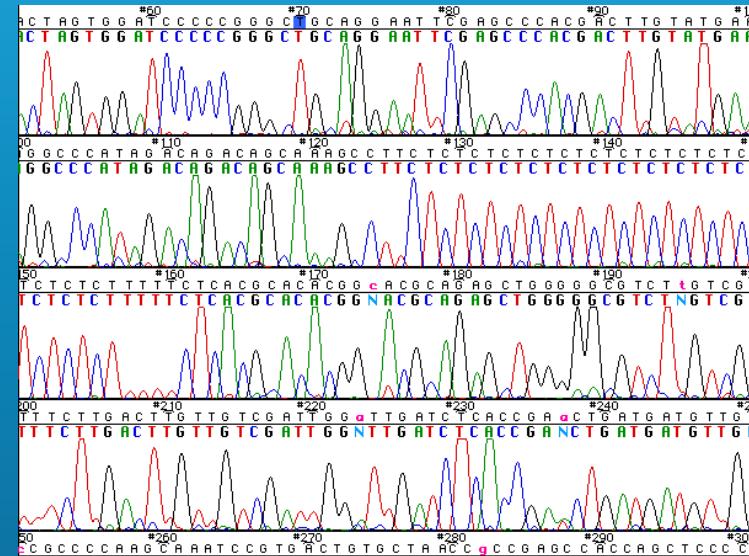


# Automatic DNA Sequencer

INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

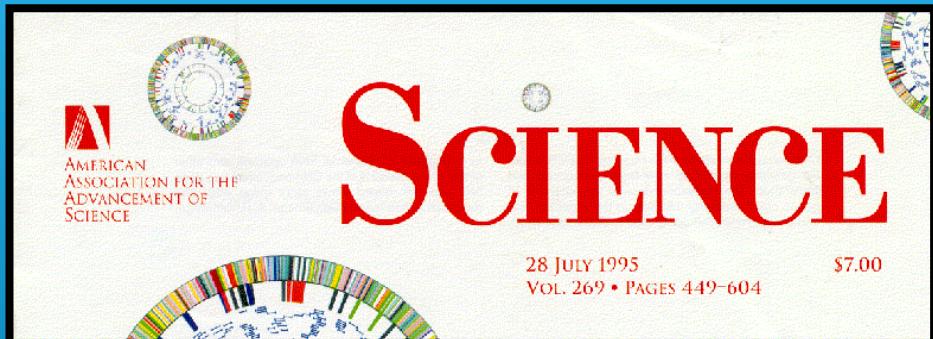
21 a 23 de Janeiro de 2020



# A Era Genômica



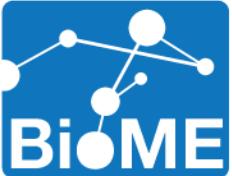
28 de Julho, 1995



## Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd

Robert D. Fleischmann, Mark D. Adams, Owen White, Rebecca A. Clayton, Ewen F. Kirkness, Anthony R. Kerlavage, Carol J. Bult, Jean-Francois Tomb, Brian A. Dougherty, Joseph M. Merrick, Keith McKenney, Granger Sutton, Will FitzHugh, Chris Fields,\* Jeannine D. Gocayne, John Scott, Robert Shirley, Li-Ing Liu, Anna Glodek, Jenny M. Kelley, Janice F. Weidman, Cheryl A. Phillips, Tracy Spriggs, Eva Hedblom, Matthew D. Cotton, Teresa R. Utterback, Michael C. Hanna, David T. Nguyen, Deborah M. Saudek, Rhonda C. Brandon, Leah D. Fine, Janice L. Fritchman, Joyce L. Fuhrmann, N. S. M. Geoghegan, Cheryl L. Gnehm, Lisa A. McDonald, Keith V. Small, Claire M. Fraser, Hamilton O. Smith, J. Craig Venter†

An approach for genome analysis based on sequencing and assembly of unselected pieces of DNA from the whole chromosome has been applied to obtain the complete nucleotide sequence (1,830,137 base pairs) of the genome from the bacterium *Haemophilus influenzae* Rd. This approach eliminates the need for initial mapping efforts and is therefore applicable to the vast array of microbial species for which genome maps are unavailable. The *H. influenzae* Rd genome sequence (Genome Sequence DataBase accession number L42023) represents the only complete genome sequence from a free-living organism.



INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020

# Human Genome

- Consórcio Internacional
  - Grupos acadêmicos
- Celera Genomics
  - Companhia Privada



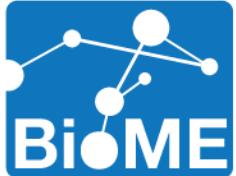
# NGS

INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020

Sequencing System	iSeq™	MiniSeq™	MiSeq®	NextSeq®	HiSeq® 4000	HiSeq® X	NovaSeq® 6000
Output per run	1.2 Gb	7.5 Gb	15 Gb	120 Gb	4000	Five/Ten	6000
Instrument price	\$19.9K	\$49.5K	\$99K	\$275K	\$900K	\$6M <sup>2</sup> /\$10M <sup>2</sup>	\$985K
Installed base <sup>3</sup>	NA	~600	~6,000	~2,400	~2,300 <sup>4</sup>	~285	



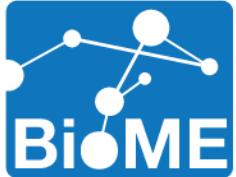
# Third Generation of Sequencers

INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020





# MGSeq200

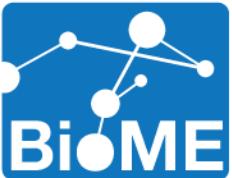
INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020



Flow cell	Read length	Data output	Q30%	TAT
FCS-100M	SE100	10G	>80%	~10h
	PE100	20G	>80%	~28h
	PE150	30G	>75%	~41h
FCL-500M	SE50	25G	>80%	~11h
	SE100	50G	>80%	~18h
	PE100	100G	>80%	~45h
	PE150	150G	>75%	~64h

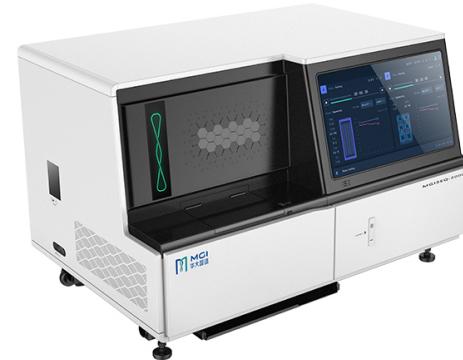


# MGI Seq2000

INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020



Model	DNBSEQ-G400	
No. of Flow Cell per run	2	
Type of Flow Cell	FCS	FCL
No. of lanes/ Flow Cell	2	4
Max reads/ Flow Cell * *	550M	1500M-1800M
Read lengths	SE100 PE100 PE150	SE50 SE100 SE400 PE100 PE150 PE200

\* The maximum number of effective reads are based on the sequencing of an internal standard library. Actual output may vary depending on sample type and library preparation method.

Read Lengths	Data Output	Data Quality Q30 *	Run Time **
SE50-FCL	75-90G	>85 %	~16h
SE100-FCL	150-180G	>85 %	~28h
SE400-FCL	600-720G	>70 %	~109h
PE100-FCL	300-360G	>85 %	~54h
PE150-FCL	450-540G	>75 %	~78h
PE200-FCL	600-720G	TBD	TBD
SE100-FCS	55G	>85 %	12-14h
PE100-FCS	110G	>85 %	~26h
PE150-FCS	165G	>75 %	~38h

\* The percentage of base above Q30 is the average of an internal standard library over the entire run. The actual performance is affected by factors such as sample type, library quality, and Insert fragment length.

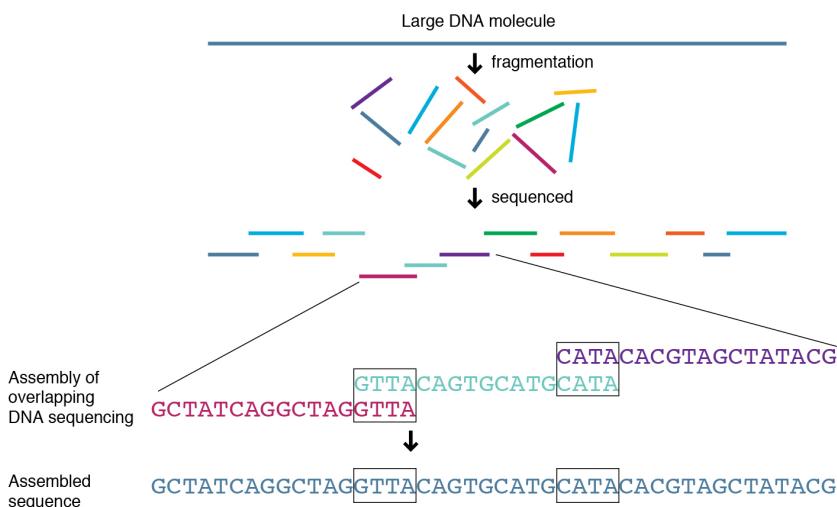
\*\* Run time was calculated based on Dual-Flow Cell mode, and includes sample loading, sequencing, base calling and data processing.

# Single-End Paired-End

INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

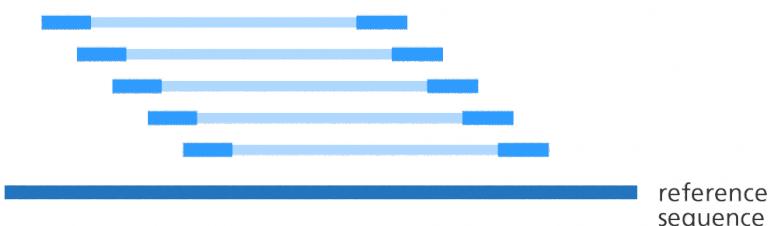
21 a 23 de Janeiro de 2020



## Single-end reads

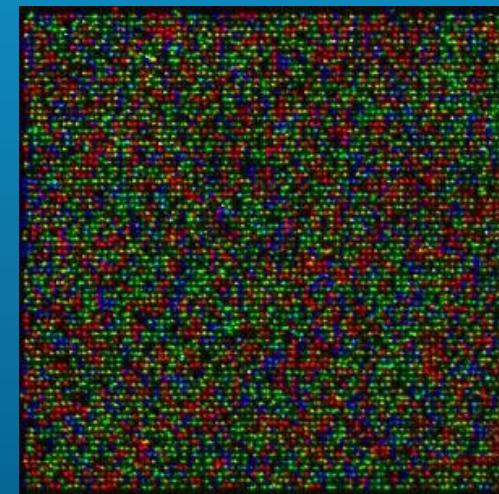
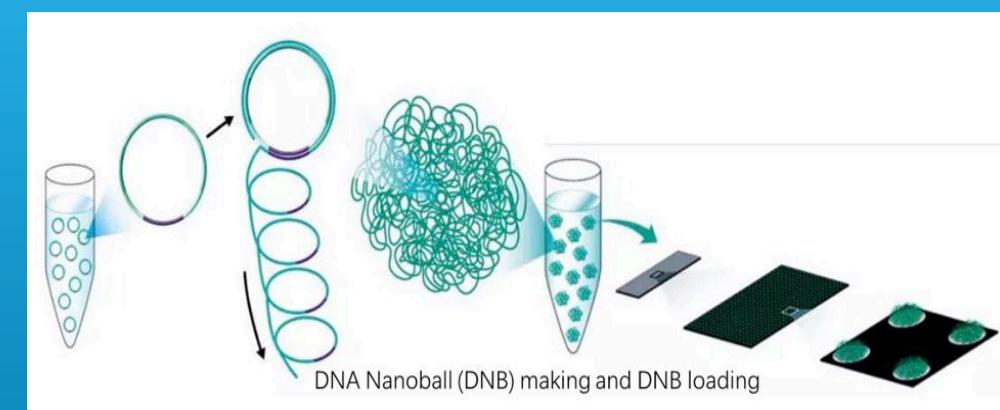
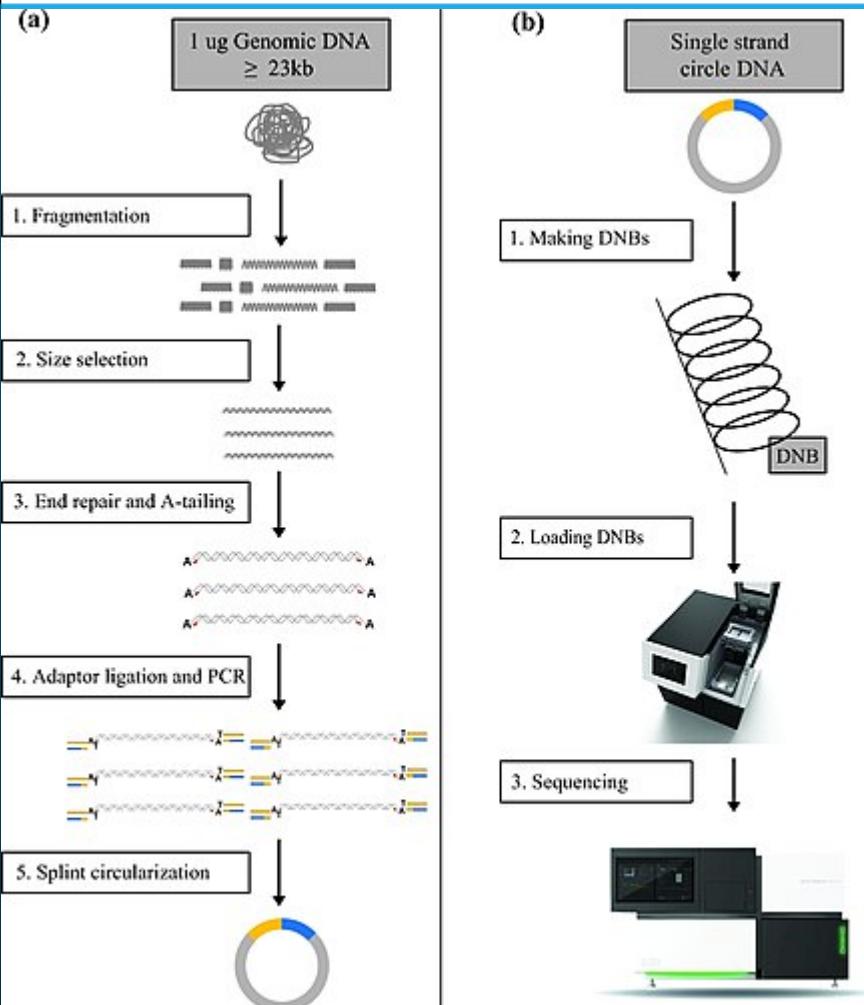


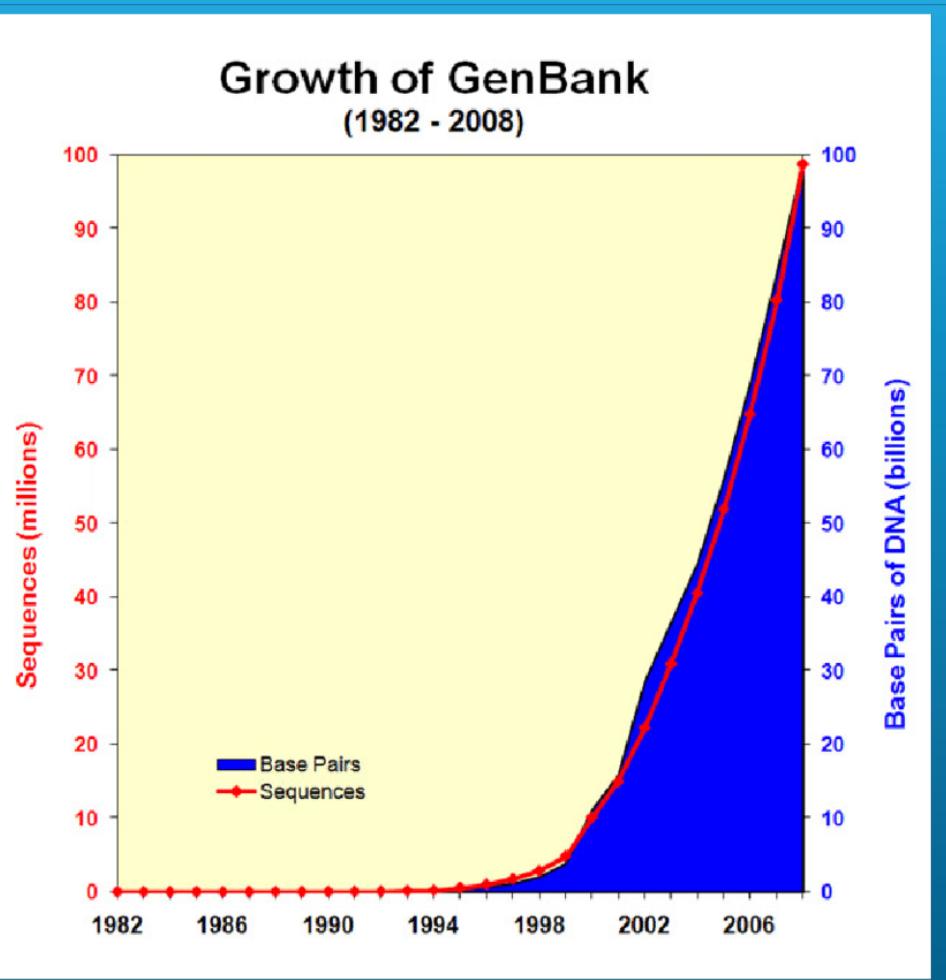
## Paired-end reads

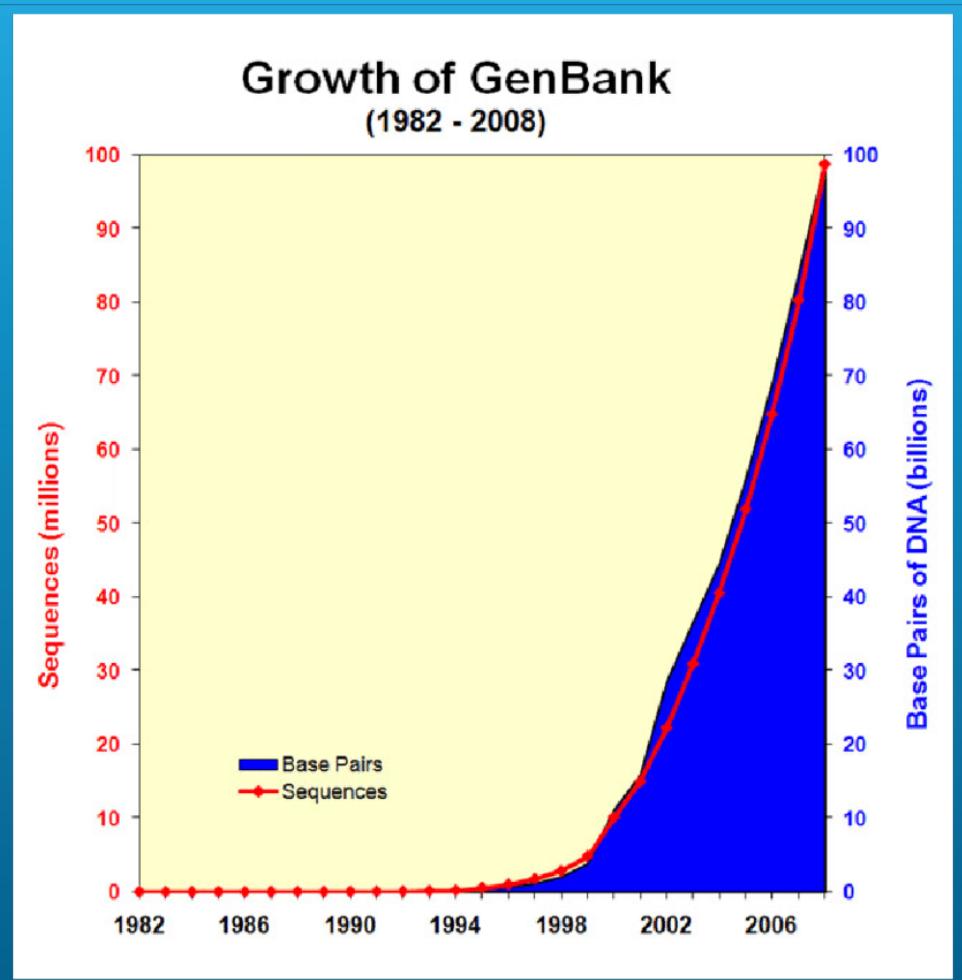


sequenced fragment      unknown sequence      sequenced fragment  
200 - 1000bp

# DNA nanoballs

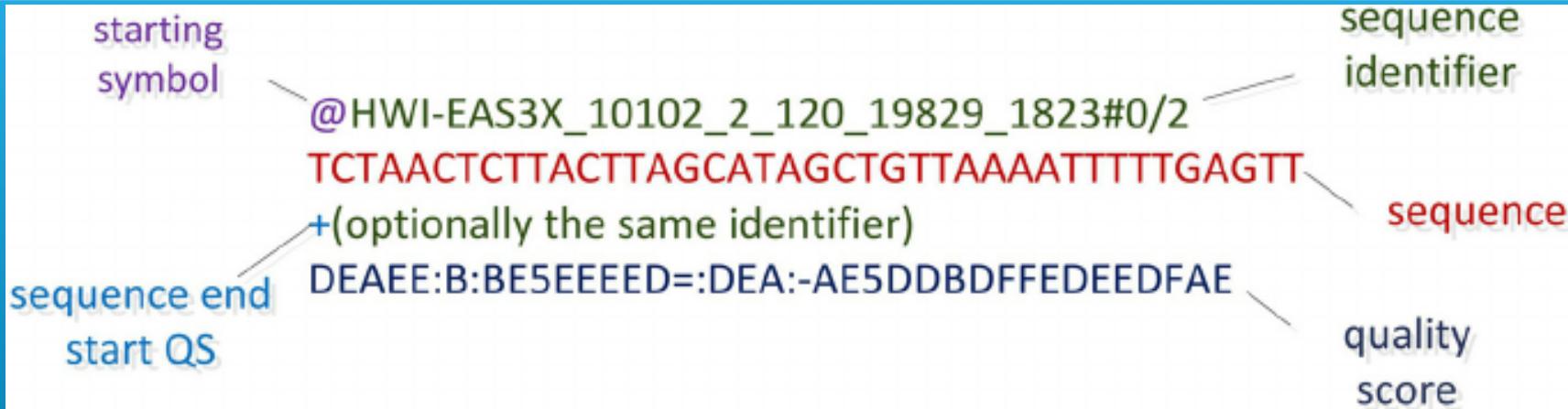






Por volta de 01  
segundo

# FASTQ Format

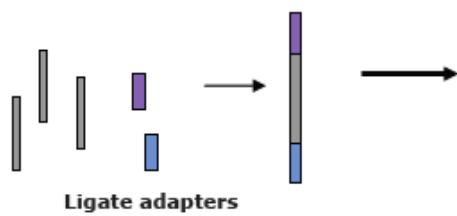


```
@HWI-M01141:63:A4NDL:1:1101:16668:1377 1:N:0:TATAGCGAGACACCGT
NACAGAGGGTCAAGCGTTAACCGATTACTGGCGTAAAGCGCGTAGGTGGTTGTTA
+
#>>AA>CAABBGGGGGGGGFFGHFEGGGGFHHHHFEGCEHHFEGGGGG@EEHHGGGHGHH
@HWI-M01141:63:A4NDL:1:1101:14849:1418 1:N:0:TATAGCGAGACACCGT
NACGAAGGGTCAAGCGTTACTCGGAATTACTGGCGTAAAGCGTGCCTAGGTGGTGGTTA
+
#>>>A??AFAA1BGGEGGAAFGGCA0BFF1D2BCF/EEG/DBEE/E?GAEEFGAEFAEFG1
@HWI-M01141:63:A4NDL:1:1101:13802:1421 1:N:0:TATAGCGAGACACCGT
NACGGAGGGTCAAGCGTTAACCGATTACTGGCGTAAAGCGCACGCAGGCCGTTGTTA
+
#>>AAABBBABBGGGGGGGG?FGHGGGGGHHHHHHHHGGGGHHGGGGGGEGGGGGEGG?FD
@HWI-M01141:63:A4NDL:1:1101:15928:1426 1:N:0:TATAGCGAGACACCGT
NACGTAGGGTGCAGCGTTAACCGATTACTGGCGTAAAGCGTGCCTAGGTGGTGGTTA
+
#>>AABFB@FBBGGGGGGGGGGHGGGGFHHHHHHHHGGGGHHGGGGGGGGGGGGEE?G4
@HWI-M01141:63:A4NDL:1:1101:14861:1431 1:N:0:TATAGCGAGACACCGT
NACGAAGGGTCAAGCGTTACTCGGAATTACTGGCGTAAAGCGTGCCTAGGTGGTGGTTA
+
#>>AAAABBFABGGGGGECEGHGGEFFHHHHHHGHGGGGHHGGGGGGEGFHGGGGGEHE
@HWI-M01141:63:A4NDL:1:1101:15264:1465 1:N:0:TATAGCGAGACACCGT
NACGTAGGGTGCAGCGTTGTCCGAATTACTGGCGTAAAGAGCTCGTAGGTGGTTGTC
+
```

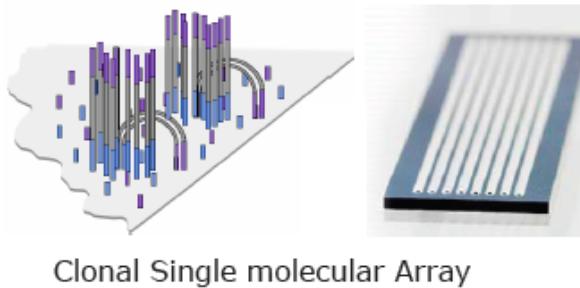
# Illumina Sequencing

## Illumina Sequencing pipeline

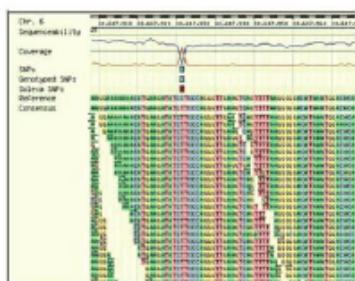
### 1. Sample Prep (1-5 days)



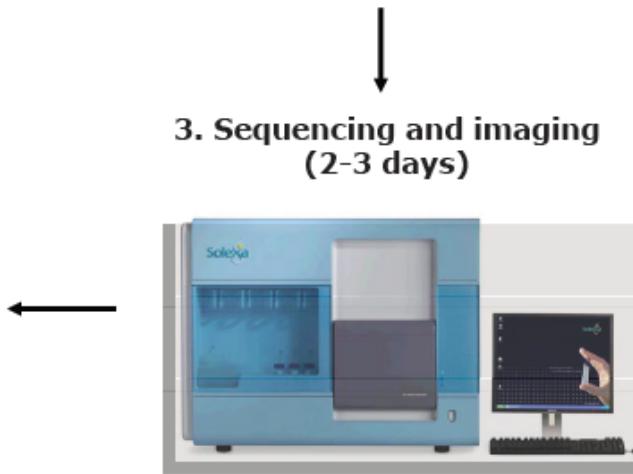
### 2. Cluster generation on flow cell (1.5 day)



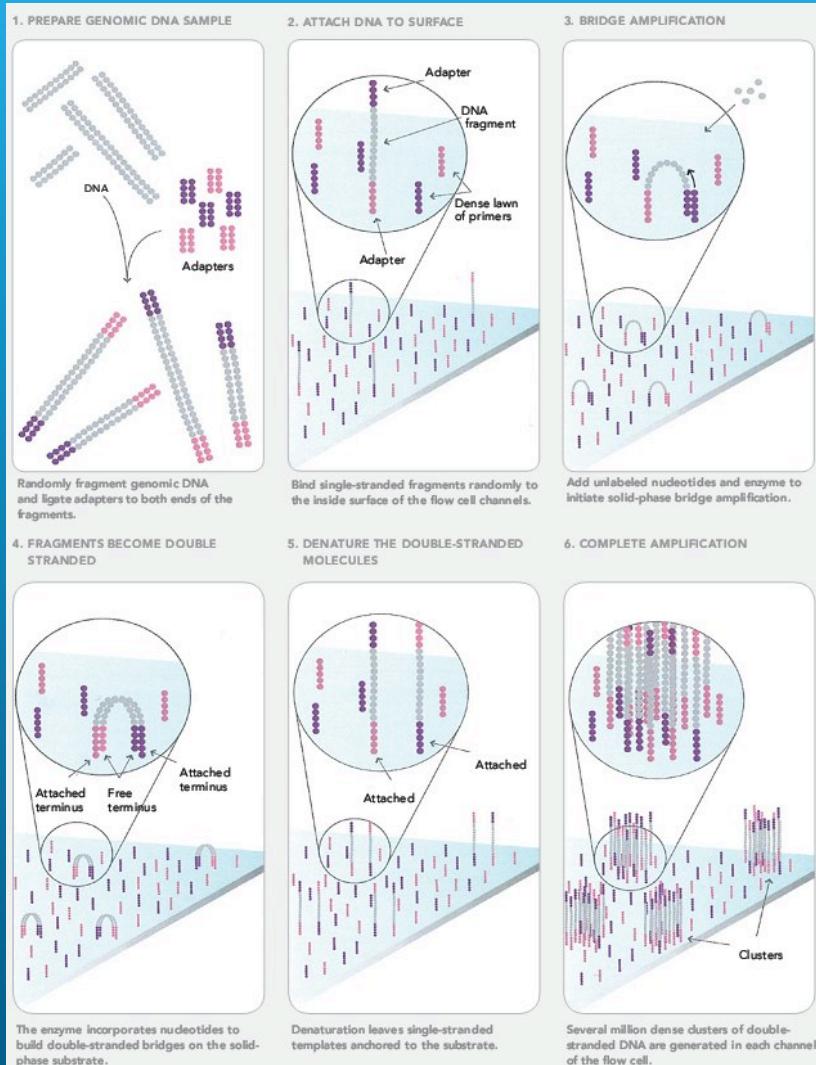
### 4. Data Analysis (days-months)



### 3. Sequencing and imaging (2-3 days)

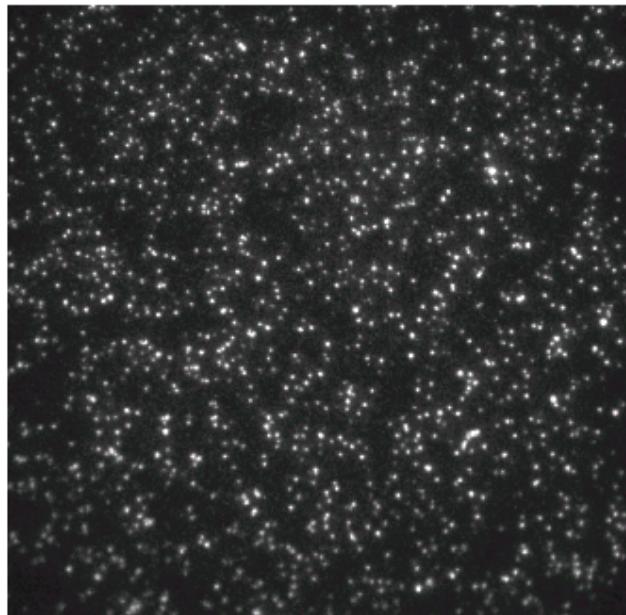


# Illumina Sequencing



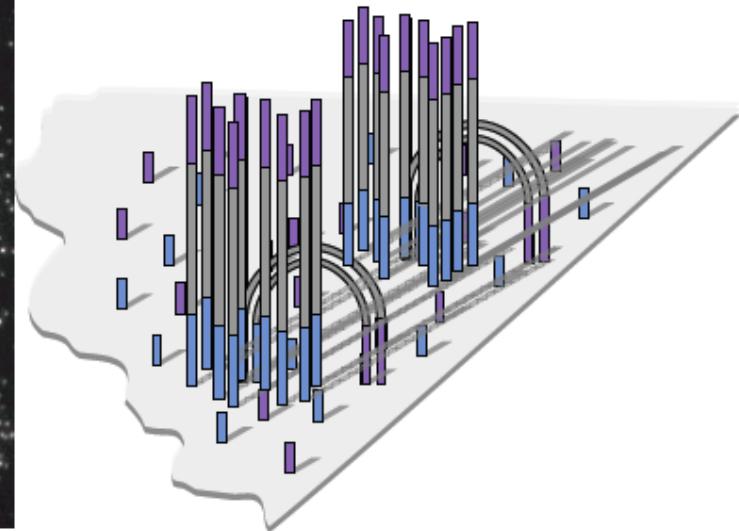
# Illumina Sequencing

## Clonal Single molecule Array



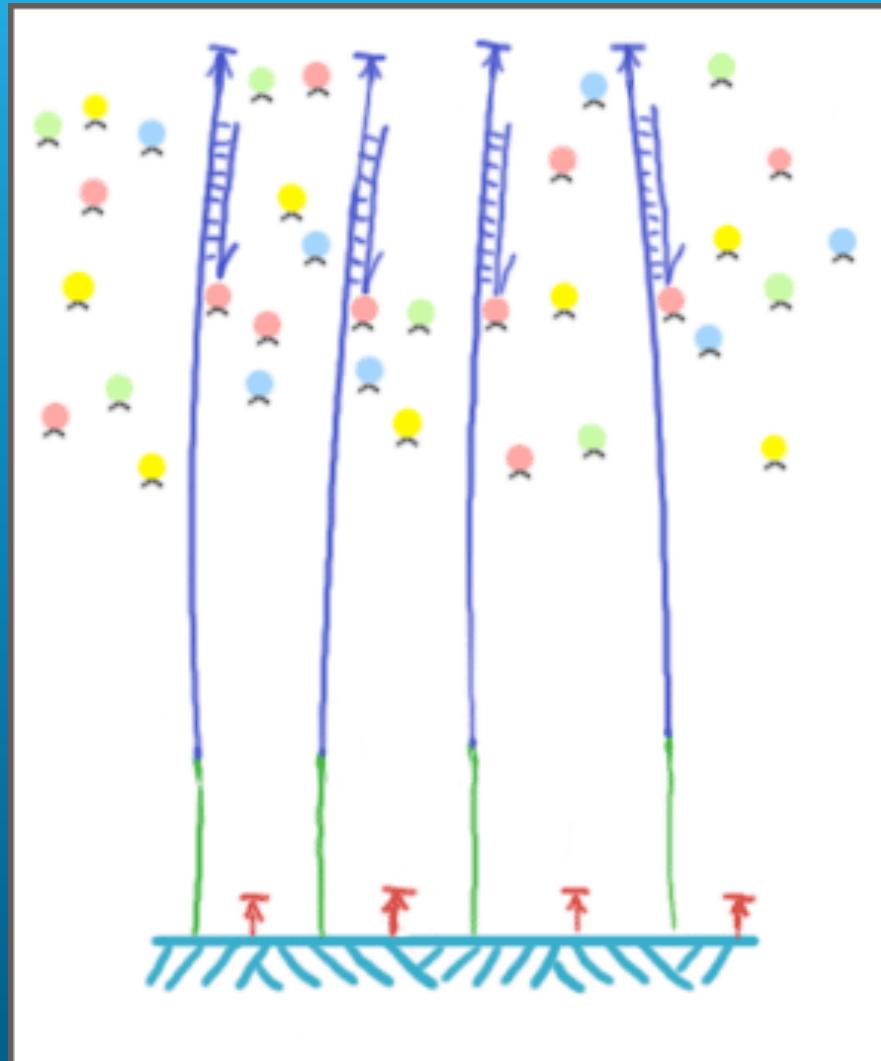
100um

Random array of clusters



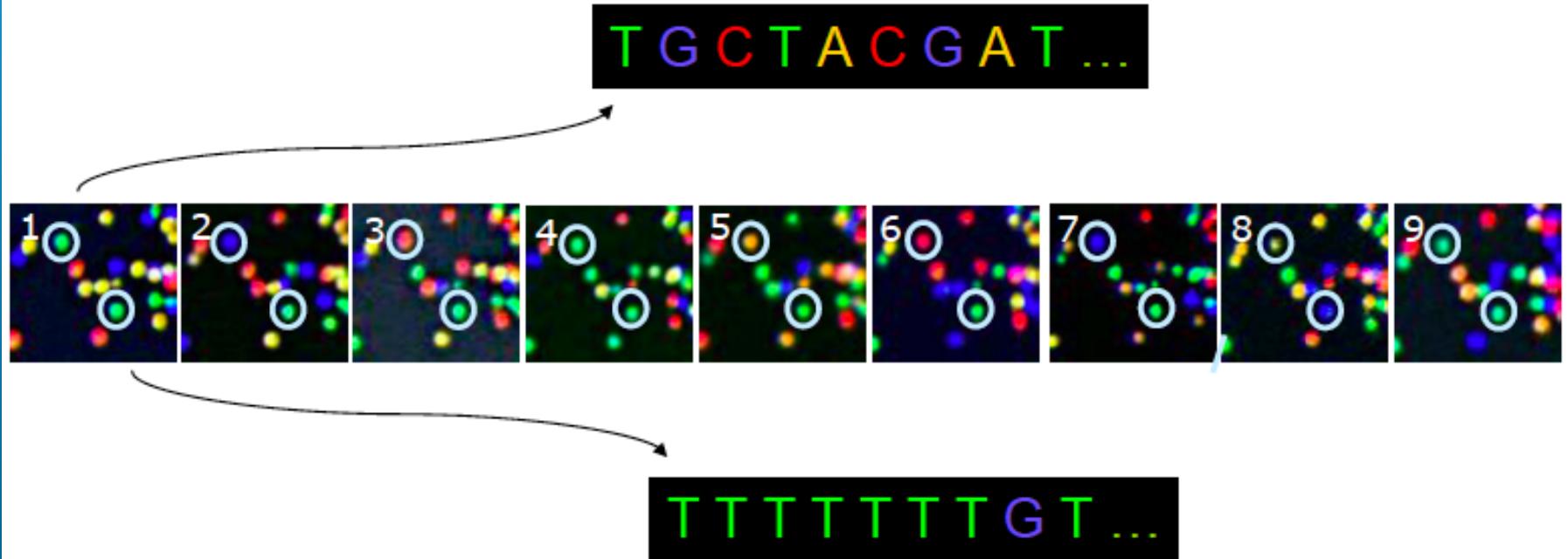
~1000 molecules per ~ 1 um cluster  
~20-30,000 clusters per tile  
~40 M clusters per flowcell

# Illumina Sequencing



# Illumina Sequencing

## Base Calling From Images

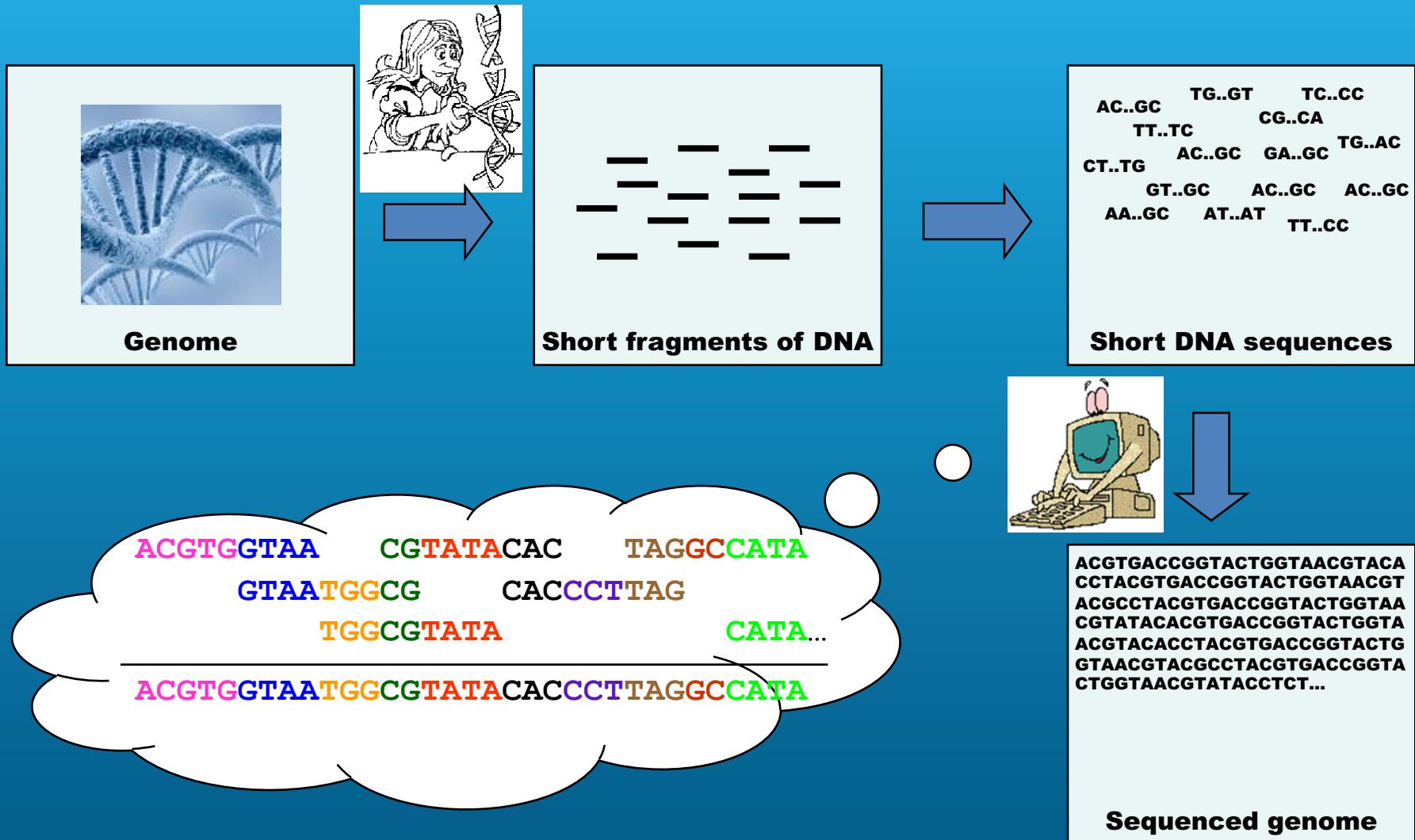


# Whole-Genome Sequencing

INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020



AAGTGCACCGAGATG

AGATGAAGTGCCCCAGT

CCAGTCAGATCGGATGC

**AAGTGCACCGAGATGAAGTGCCCCAGTCAGATCCGATGC**

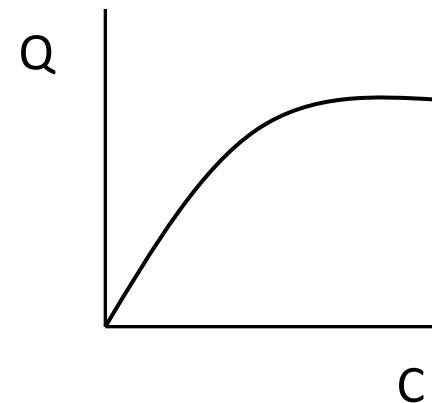
**Consenso (contig) gerado a partir de alinhamentos entre todas as sequências geradas.**

O método “shotgun do genoma inteiro” exige uma alta redundância entre as sequências geradas, o que define o conceito de cobertura.

**Genoma= 5MB**

**Sequências geradas = 50MB**

**Cobertura = 10X**

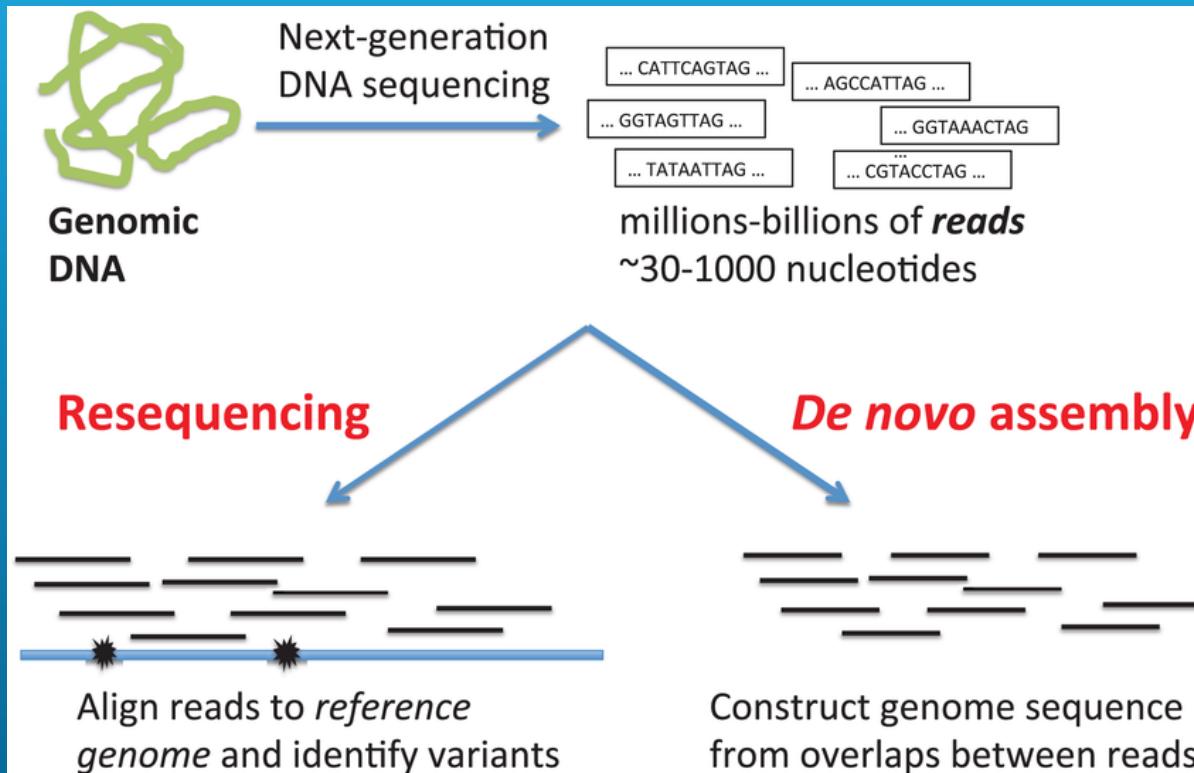


# *De novo /* resequencing

INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020

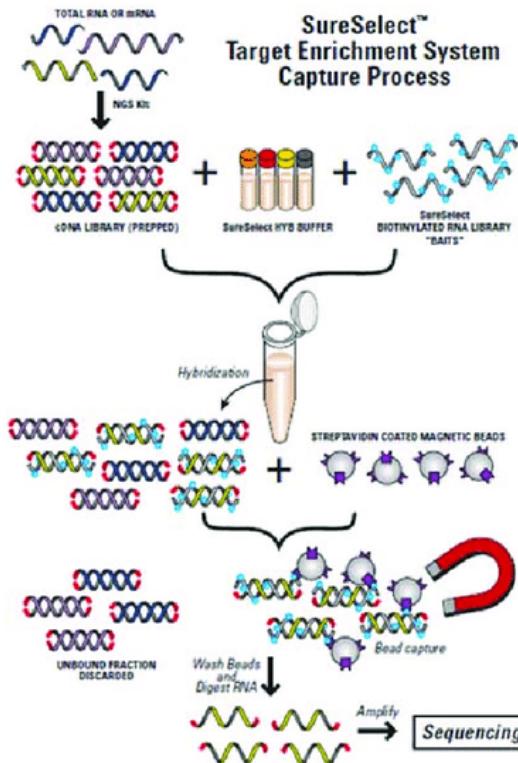
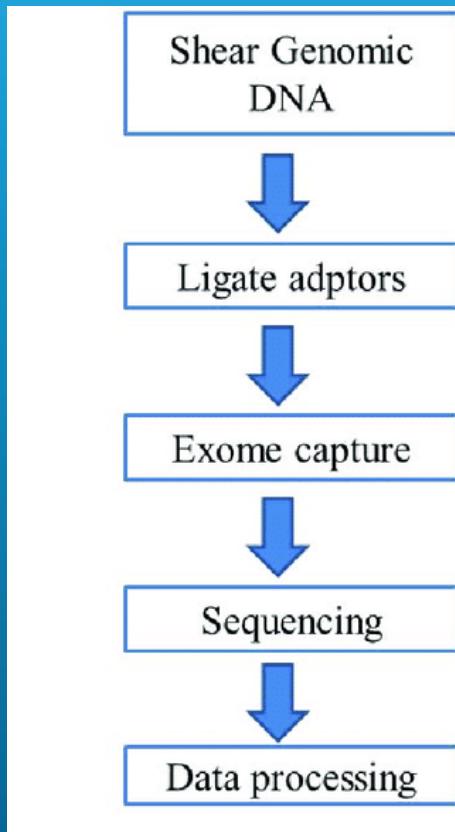


# Exome

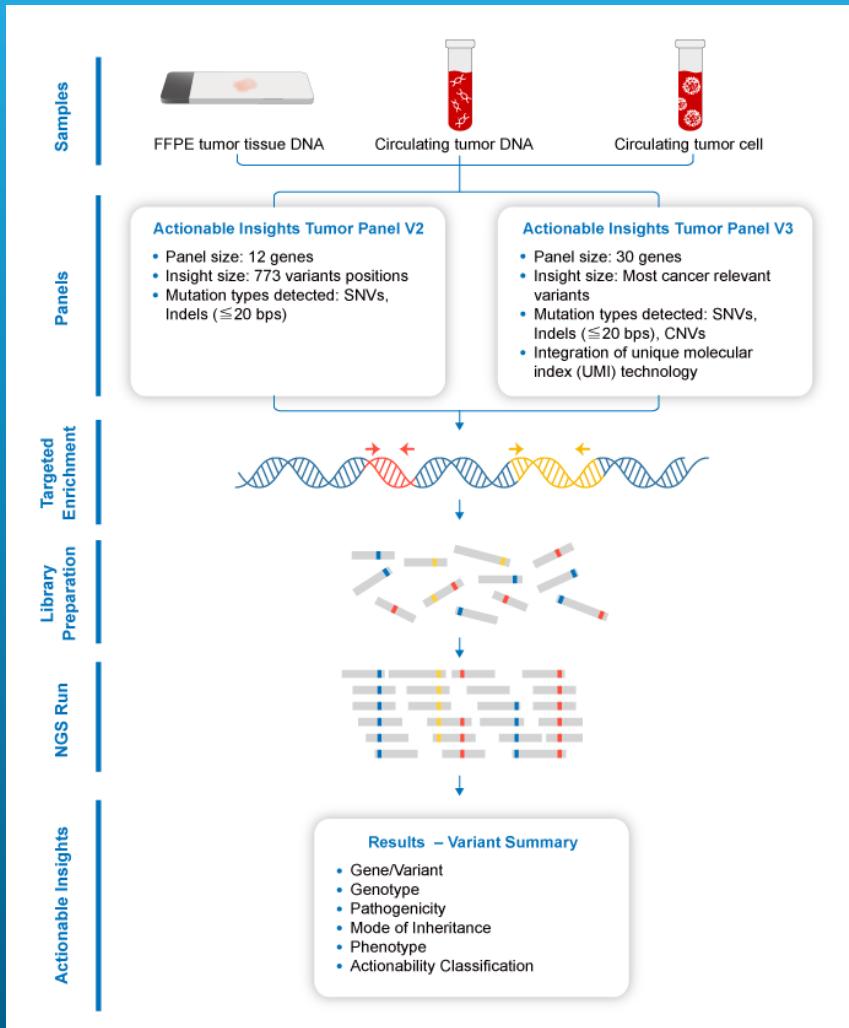
INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

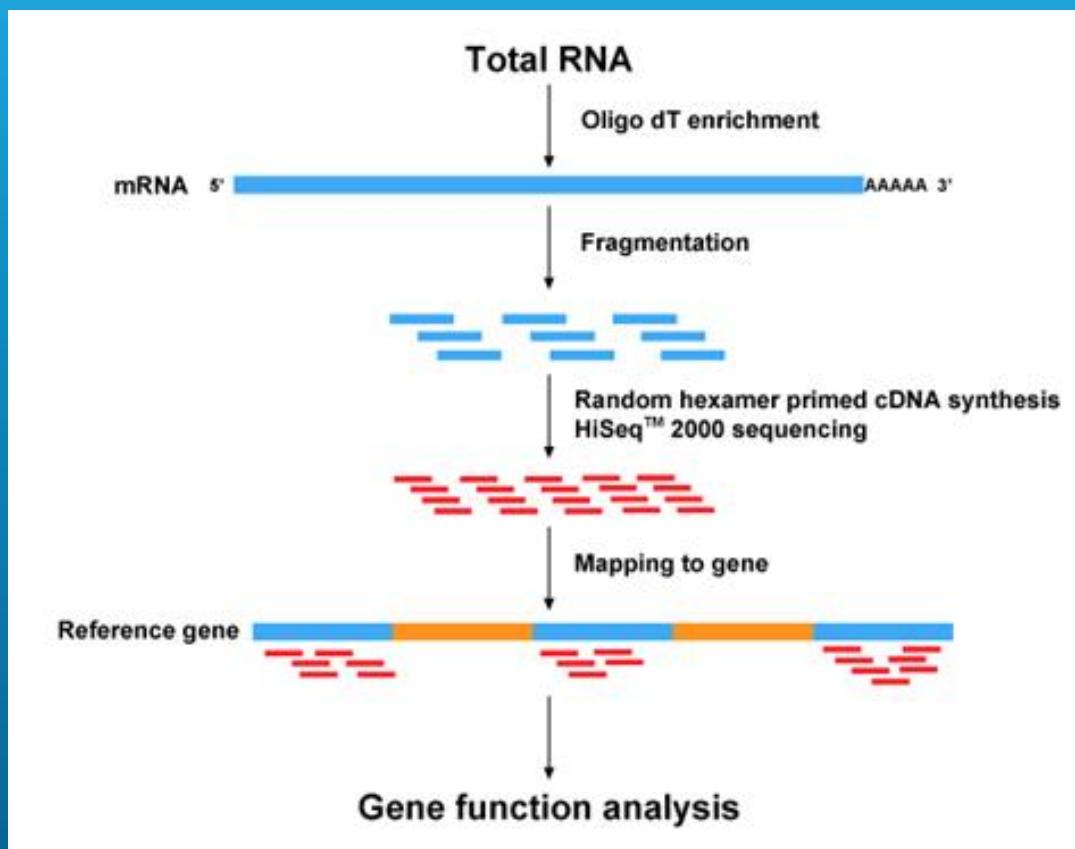
21 a 23 de Janeiro de 2020



# Gene Panel

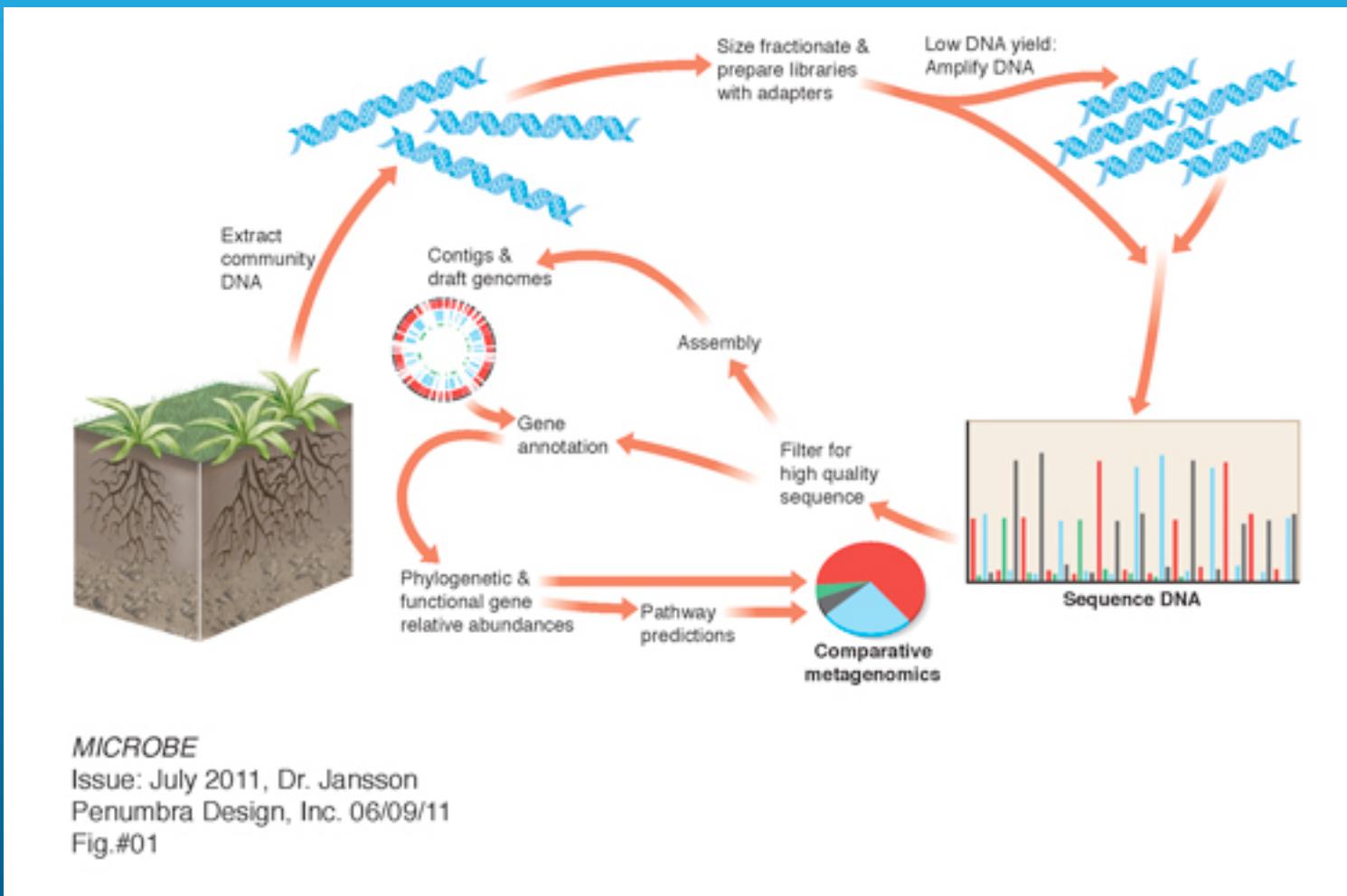


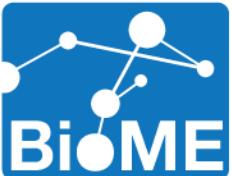
# RNA-Seq



✓ mRNAs

✓ Small RNAs





# PathoMAP

INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020

Patho **M** **A** **P**

HOME ABOUT PEOPLE DATA MAP BLOG PRESS LINKS CONTACT US

## Creating a Molecular Portrait of New York City

One swab at a time

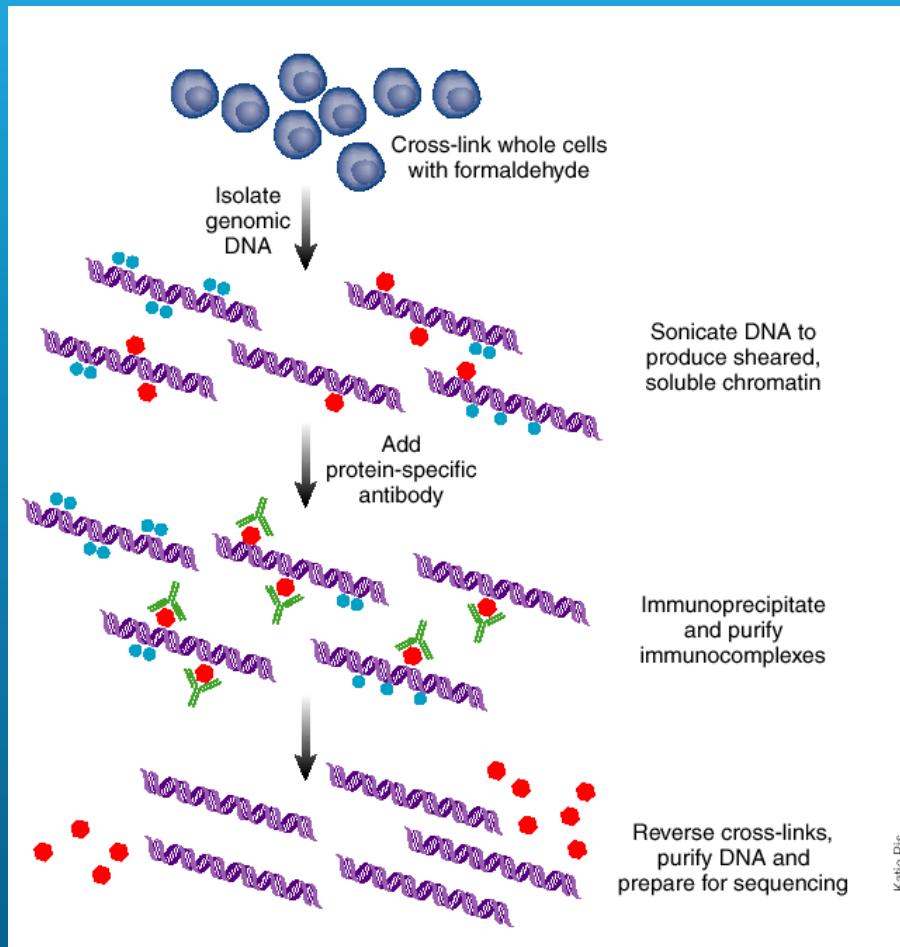
PathoMap is a research project by Weill Cornell Medical College to study the microbiome and metagenome of the built environment of NYC.

Check out the full manuscript published at Cell Systems [here](#).

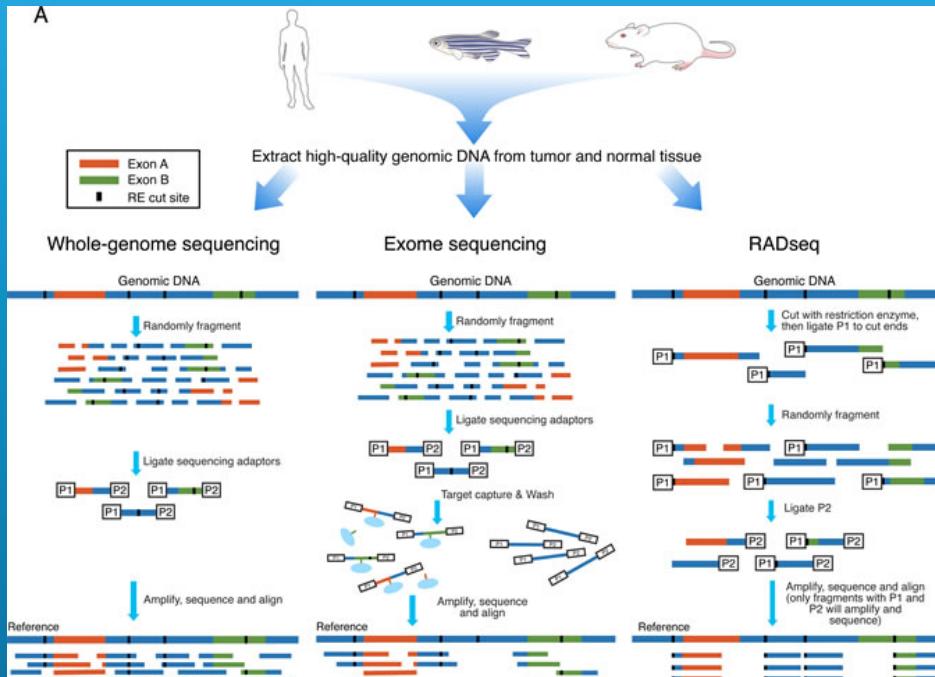
Raw data is now online [here](#).

Check out the PathoMap Ancestry Pipeline now set up on the Arvados platform [here](#)!

# ChIP-Seq



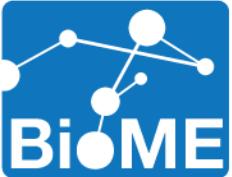
# RAD-Seq



**B**

	WGS	WES	RAD
Regions of genome targeted	All accessible* regions of the genome; genes and intergenic	Representation determined by the capture kit; enriching for exons and UTRs	Largely unbiased representation of genome; genes and intergenic
Percent genome represented	Fixed, typically ~90% genome	Fixed, typically ~2% genome	Flexible, typically 0.1%~10% genome
Species specific reagents	Same reagents can be used across species	Species-specific capture step	Same reagents can be used across species

Perry et al., 2017

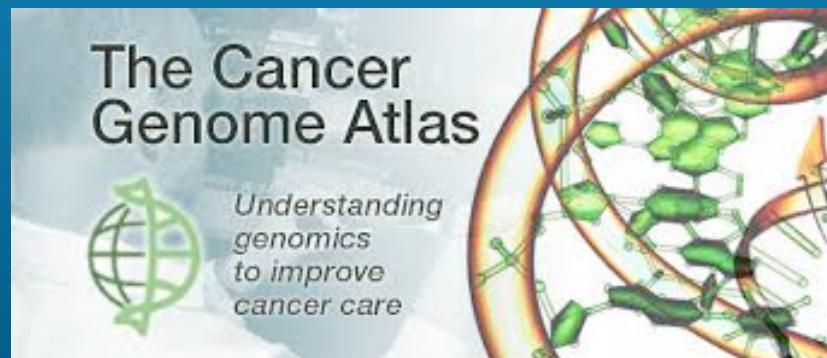
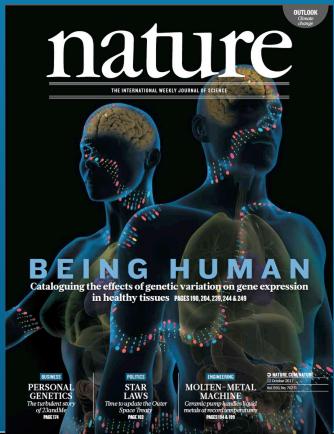
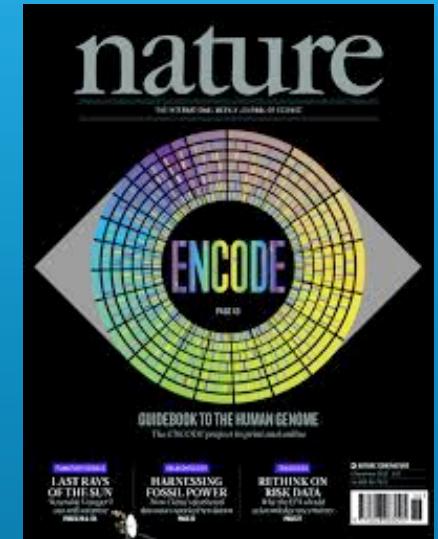


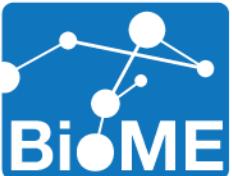
# NGS - Applications

INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020





# NGS - Applications

INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020

The NEW ENGLAND JOURNAL of MEDICINE

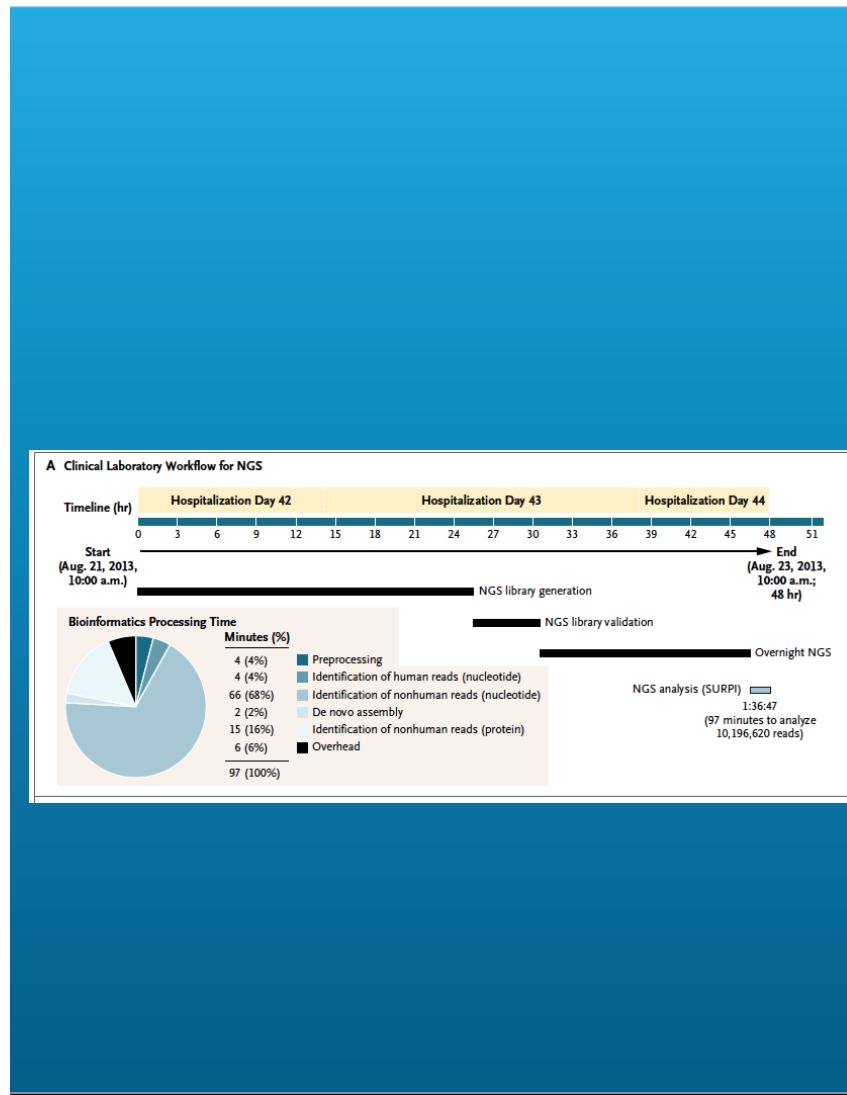
BRIEF REPORT

## Actionable Diagnosis of Neuroleptospirosis by Next-Generation Sequencing

Michael R. Wilson, M.D., Samia N. Naccache, Ph.D., Erik Samayoa, B.S., C.L.S.,  
Mark Biagtan, M.D., Hiba Bashir, M.D., Guixia Yu, B.S.,  
Shahriar M. Salarnat, M.D., Ph.D., Sneha Somasekar, B.S., Scot Federman, B.A.,  
Steve Miller, M.D., Ph.D., Robert Sokolic, M.D., Elizabeth Garabedian, R.N., M.S.L.S.,  
Fabio Candotti, M.D., Rebecca H. Buckley, M.D., Kurt D. Reed, M.D.,  
Teresa L. Meyer, R.N., M.S., Christine M. Seroogy, M.D., Renee Galloway, M.P.H.,  
Sheryl L. Henderson, M.D., Ph.D., James E. Gern, M.D., Joseph L. DeRisi, Ph.D.,  
and Charles Y. Chiu, M.D., Ph.D.

### SUMMARY

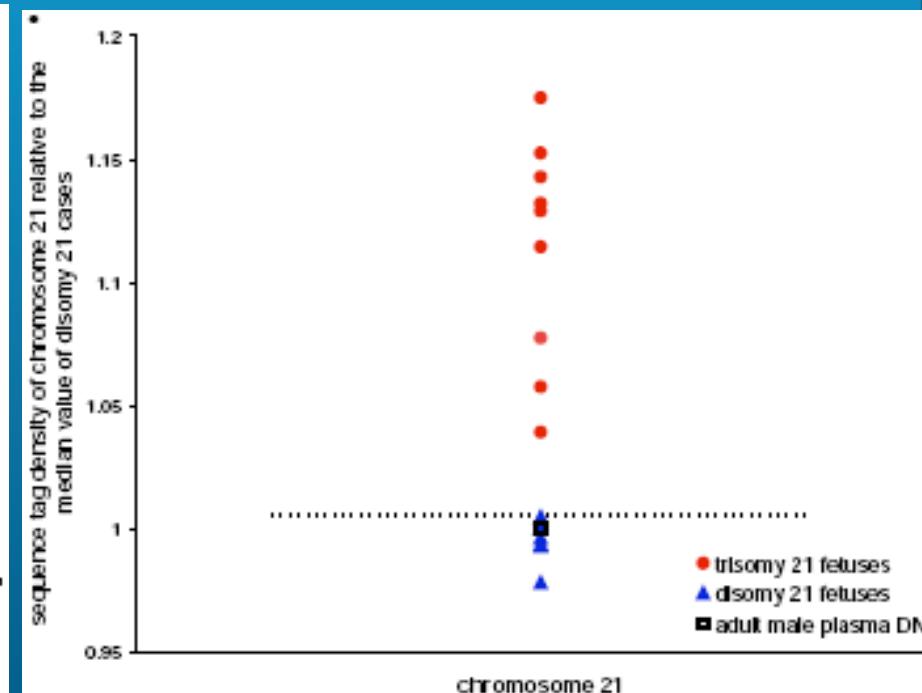
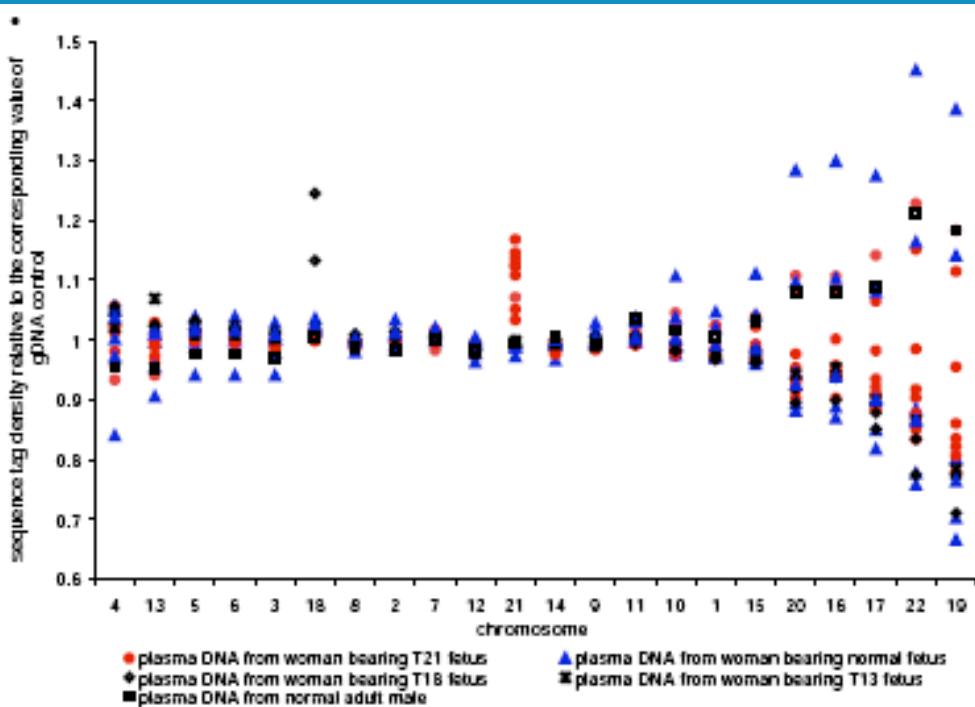
A 14-year-old boy with severe combined immunodeficiency presented three times to a medical facility over a period of 4 months with fever and headache that progressed to hydrocephalus and status epilepticus necessitating a medically induced coma. Diagnostic workup including brain biopsy was unrevealing. Unbiased next-generation sequencing of the cerebrospinal fluid identified 475 of 3,063,784 sequence reads (0.016%) corresponding to leptospira infection. Clinical assays for leptospirosis were negative. Targeted antimicrobial agents were administered, and the patient was discharged home 32 days later with a status close to his premorbid condition. Polymerase-chain-reaction (PCR) and serologic testing at the Centers for Disease Control and Prevention (CDC) subsequently confirmed evidence of *Leptospira santarosai* infection.



# NGS - Applications

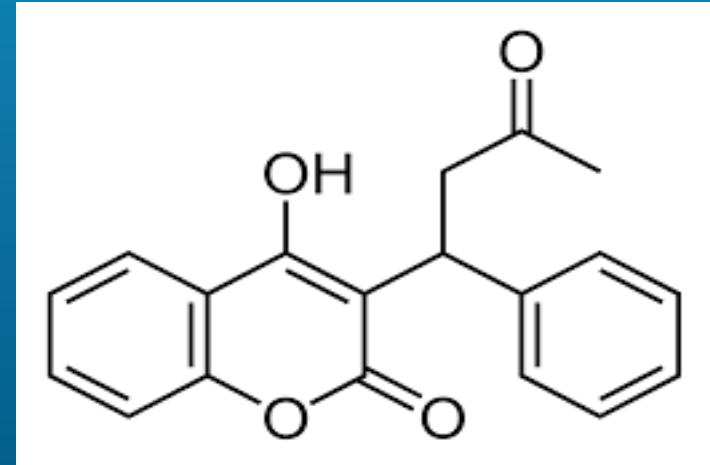
## Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood

H. Christina Fan\*, Yair J. Blumenfeld†, Usha Chitkara†, Louanne Hudgins‡, and Stephen R. Quake \*§



# Pharmacogenomics

GCCCTTCAGCCTCTAACAGTACTTAAACCAATTAA  
AAGACTCCTGTTAGTTACCTCCCCACATCCCCACC  
CGCAGGACGCTCCGTGAT**GAGCAGCTAGCTGGC**  
TGTCAGCTGTGGATCACCAAGATTGCATGGAG  
TGGGGCTGAGCTGACCAAGGGGGATGAGGGGC

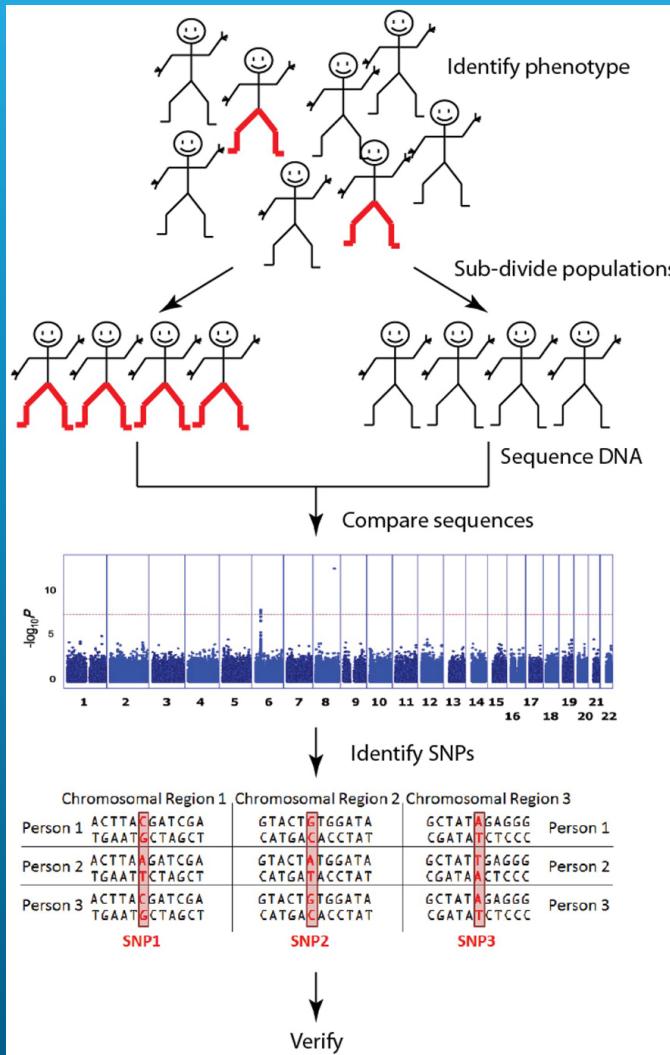


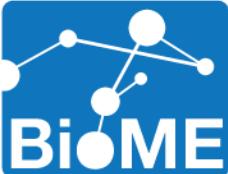
# GWAS

INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020





# NGS - Applications

## Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus

Laura S. Weyrich<sup>1</sup>, Sebastian Duchene<sup>2</sup>, Julien Soubrier<sup>1</sup>, Luis Arriola<sup>1</sup>, Bastien Llamas<sup>1</sup>, James Breen<sup>1</sup>, Alan G. Morris<sup>3</sup>, Kurt W. Alt<sup>4,5,6,7</sup>, David Caramelli<sup>8</sup>, Veit Dresely<sup>5,6</sup>, Milly Farrell<sup>9</sup>, Andrew G. Farrer<sup>1</sup>, Michael Francken<sup>10</sup>, Neville Gully<sup>11</sup>, Wolfgang Haak<sup>1</sup>, Karen Hardy<sup>12,13</sup>, Katerina Harvati<sup>10</sup>, Petra Held<sup>14</sup>, Edward C. Holmes<sup>2</sup>, John Kaidonis<sup>11</sup>, Carles Lalueza-Fox<sup>15</sup>, Marco de la Rasilla<sup>16</sup>, Antonio Rosas<sup>17</sup>, Patrick Semal<sup>18</sup>, Arkadiusz Soltysiak<sup>19</sup>, Grant Townsend<sup>11</sup>, Donatella Usai<sup>20</sup>, Joachim Wahl<sup>21</sup>, Daniel H. Huson<sup>22</sup>, Keith Dobney<sup>23,24,25</sup> & Alan Cooper<sup>1</sup>

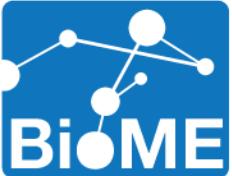
Recent genomic data have revealed multiple interactions between Neanderthals and modern humans<sup>1</sup>, but there is currently little genetic evidence regarding Neanderthal behaviour, diet, or disease. Here we describe the shotgun-sequencing of ancient DNA from five specimens of Neanderthal calcified dental plaque (calculus) and the characterization of regional differences in Neanderthal ecology. At Spy cave, Belgium, Neanderthal diet was heavily meat based and included woolly rhinoceros and wild sheep (mouflon), characteristic of a steppe environment. In contrast, no meat was detected in the diet of Neanderthals from El Sidrón cave, Spain, and dietary components of mushrooms, pine nuts, and moss reflected forest gathering<sup>2,3</sup>. Differences in diet were also linked to an overall shift in the oral bacterial community (microbiota) and suggested that meat consumption contributed to substantial variation within Neanderthal microbiota. Evidence for self-medication was detected in an El Sidrón Neanderthal with a dental abscess<sup>4</sup> and a chronic gastrointestinal pathogen (*Enterocytozoon bieneusi*). Metagenomic data from this individual also contained a nearly complete genome of the archaeal commensal *Methanobrevibacter oralis* (10.2× depth of coverage)—the oldest draft microbial genome generated to date, at around 48,000 years old. DNA preserved within dental calculus represents a notable source of information about the behaviour and health of ancient hominin specimens, as well as a unique system that is useful for the study of long-term microbial evolution.

Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus

purposes<sup>8</sup>. As a result, Neanderthal diet remains a topic of considerable debate, with limited data on the specific animals and plants directly consumed or the potential effects on Neanderthal health and disease.

Although genomic studies continue to reveal evidence of interbreeding between anatomically modern humans and Neanderthals across Eurasia<sup>9</sup>, little is known about the health consequences of these interactions. The genetic analysis of Neanderthal dental calculus represents an opportunity to examine this issue and to reconstruct Neanderthal diet, behaviour, and disease<sup>10</sup>. Here, we report the first genetic analysis of dental calculus from five Neanderthals (two individuals from El Sidrón cave in Spain; two individuals from Spy cave in Belgium; and one individual from Breuil Grotta in Italy) and compare these data to a historic wild-caught chimpanzee ( $n = 1$ ) and modern human ( $n = 1$ ), as well as to low coverage sequencing of calculus from a wide-range of ancient humans (Supplementary Table 1). To provide increased resolution of the diseases that may have affected Neanderthals, we also deeply sequenced (>147 million reads) dental calculus from the best-preserved Neanderthal, El Sidrón 1, which suffered from a dental abscess<sup>4</sup>.

Size-based PCR-amplification biases can confound standard metabarcoding analyses (for example, sequencing of 16S ribosomal (r)RNA amplicons<sup>11,12</sup>) of ancient dental calculus<sup>13</sup>. Consequently, we compared metagenomic-shotgun sequencing and 16S rRNA amplicon (V4 region) analyses of the Neanderthal dental calculus specimens—by far the oldest specimens to date. The 16S amplicon data were used to validate



# Bioinformática

INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

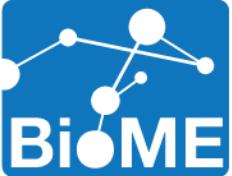
21 a 23 de Janeiro de 2020

## O Bioinformata

- Manipula a informação.
- Desenvolve ferramentas
- Bancos de dados locais.
- Local.
- Muita programação.
- Habilidades de TI.

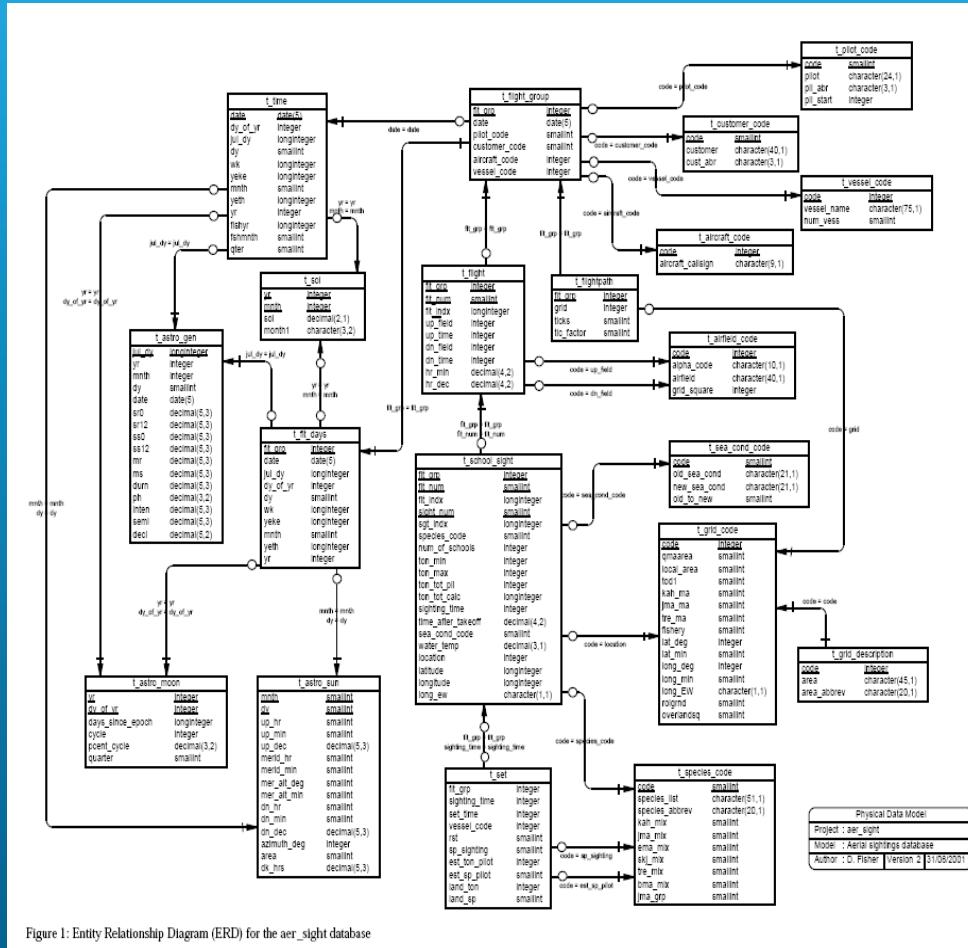
## O Usuário

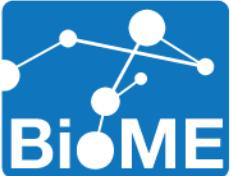
- Recursos da Web.
- Local ou remoto.
- Nada de programação.
- Pouca habilidade de TI.



- Familiarizar-se com Linux
- Aprender a usar bancos relacionais
- Familiarizar-se com R
- Aprender a programar (PERL)

# Bancos Relacionais





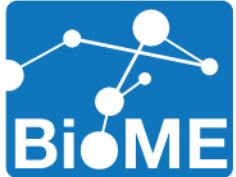
# Bancos Relacionais

INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020

- Mais fácil de aprender do que programação.
- Ideal para o estudante “padrão”.
- Estão disponíveis em várias projetos.

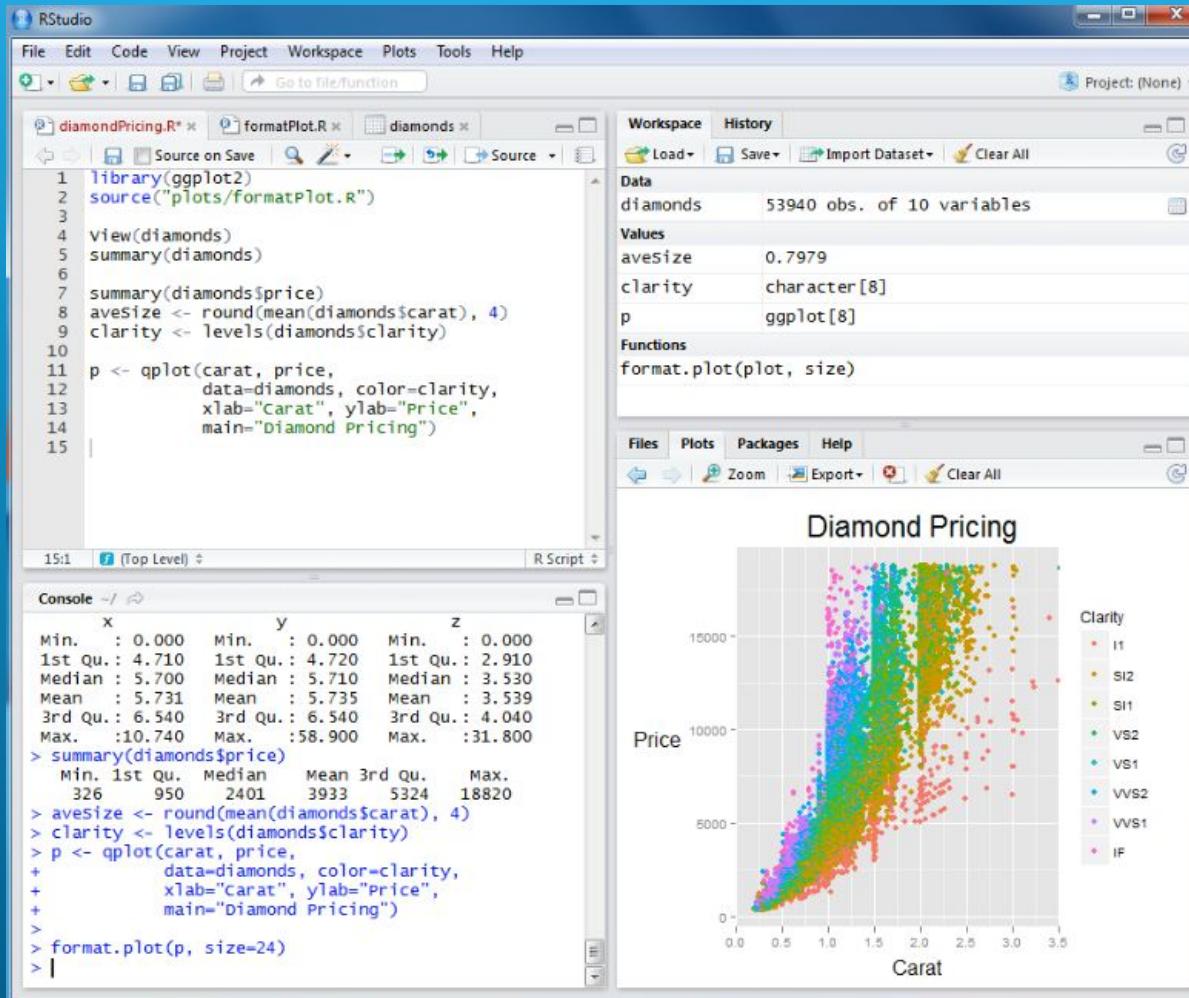


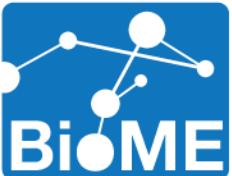
# R Language

INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020





# O meu genoma

INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020

**FOLHA DE S.PAULO**

LOGIN ASSINE A FOLHA ATENDIMENTO

SÃO PAULO 10.8°C OUTRAS CIDADES

UM JORNAL A SERVIÇO DO BRASIL

QUINTA-FEIRA, 15 DE AGOSTO DE 2013 22H13

Site OK

TEMAS DO DIA MENSALÃO · CELULAR PRÉ-PAGO · DÓLAR · EGITO · FUTEBOL CLASSIFICADOS TV FOLHA HORÓSCOPO ACERVO FOLHA

ÚLTIMAS NOTÍCIAS Megainvestidor Buffett vende US\$ 38 mi em ações da dona do 'USA Today'

**opinião**

Maior | Menor Enviar por e-mail Comunicar erros Link <http://folha.com/no> +

24/03/2013 - 03h11

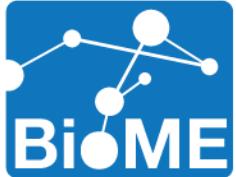
## Sandro José de Souza: O meu genoma

Recomendar 93 +1 0 Ouvir o texto

**TENDÊNCIAS / DEBATES**

Recentemente, tornei acessível a quem interessar a sequência do meu próprio genoma. Ela está disponível no banco público mantido pelo Instituto Nacional de Saúde dos EUA ([www.ncbi.nlm.nih.gov/sra/SRX208349](http://www.ncbi.nlm.nih.gov/sra/SRX208349)), bem como nas páginas do Instituto do Cérebro da Universidade Federal do Rio Grande do Norte e do Instituto de Bioinformática e Biotecnologia.

**FOLHA** Folha de S.Paulo Curtir 1.808.048 pessoas curtiram Folha de S.Paulo.



# Artigo de Marcelo Leite para a FSP

INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020



The screenshot shows a news article from Folha de S. Paulo's ilustríssima section. The header includes the newspaper's name, a three-star rating, and navigation links for menu, subscription, and login. The main title of the article is "Conhecer o próprio genoma envolve surpresas e decepções". A subtitle below it reads: "Poder preditivo dos genes sobre pessoas sadias ainda é limitado, mas não irrelevante".

MENU ASSINE

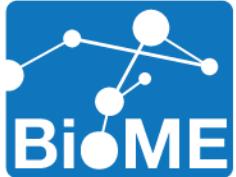
FOLHA DE S.PAULO

ilustríssima > clube de leitura folha cartuns capas ilustríssima conversa

Conhecer o próprio genoma envolve surpresas e decepções

Poder preditivo dos genes sobre pessoas sadias ainda é limitado, mas não irrelevante

<https://www1.folha.uol.com.br/ilustrissima/2019/03/conhecer-o-proprio-genoma-envolve-surpresas-e-decepcoes.shtml>



INTRODUÇÃO À ANÁLISE DE  
DADOS DE SEQUENCIADORES DE  
SEGUNDA GERAÇÃO

Fiocruz/Biomanguinhos  
Rio de Janeiro - RJ

21 a 23 de Janeiro de 2020

[sandro@neuro.ufrn.br](mailto:sandro@neuro.ufrn.br)

[sandro@sandrodesouza.com.br](mailto:sandro@sandrodesouza.com.br)