



---

**Dando boas vindas!**

Marcel Caraciolo, CTO  
[marcel@genomika.com.br](mailto:marcel@genomika.com.br)



# Quem somos ?

- › Um dos laboratórios mais avançados em clínica genética do Brasil e o primeiro localizado na região Norte e Nordeste.
  
- › Portfólio de testes genéticos para diagnóstico e tratamento personalizado de doenças hereditárias, raras e tumores.





# Sequenciamento Clínico



Mercado mundial em crescimento e recente no Brasil

**EXAME**

## O EFEITO ANGELINA

O sequenciamento genético, que permitiu à atriz Angelina Jolie se livrar do risco de ter um câncer de mama, começa a se popularizar no Brasil graças a uma nova geração de startups

GUILHERME MANEIRINI

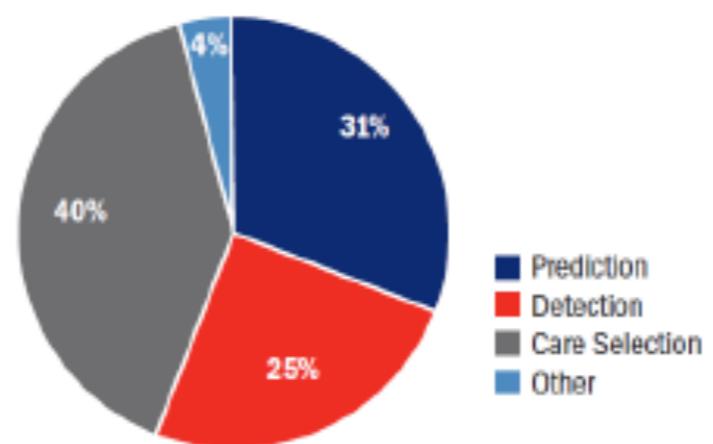


### Breast cancer: What is the risk?

By Ben Carter  
BBC News



- Tests and Genomic Applications by Clinical Impact Area, BoozAH, 2013



Source: Genetic Tests and Genomic Applications In Practice and Prevention Database  
(N= 450), CDC Office of Public Health Genomics 2012



**MERCK**



ALEXION

pharmaceuticals





# Tecnologia com especialistas



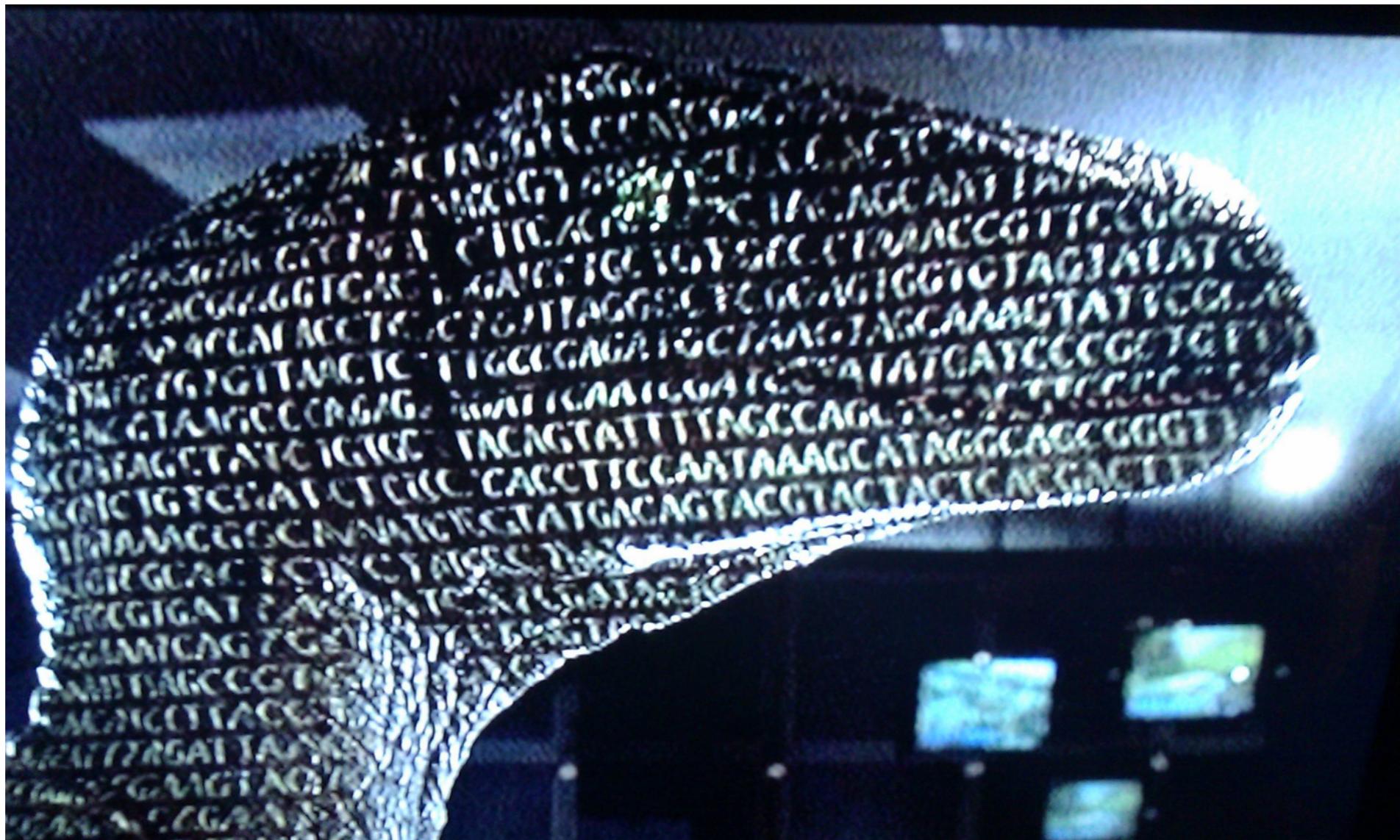
Fusão de especialistas em biologia molecular e tecnologia de informação



João Bosco Oliveira, CEO e Co-Fundador  
M.D., PhD e ex-chefe de pesquisa do serviço  
de imunologia e genética do Dpto. de Medicina  
Laboratorial, Centro Clínico, National Institutes  
of Health , USA.

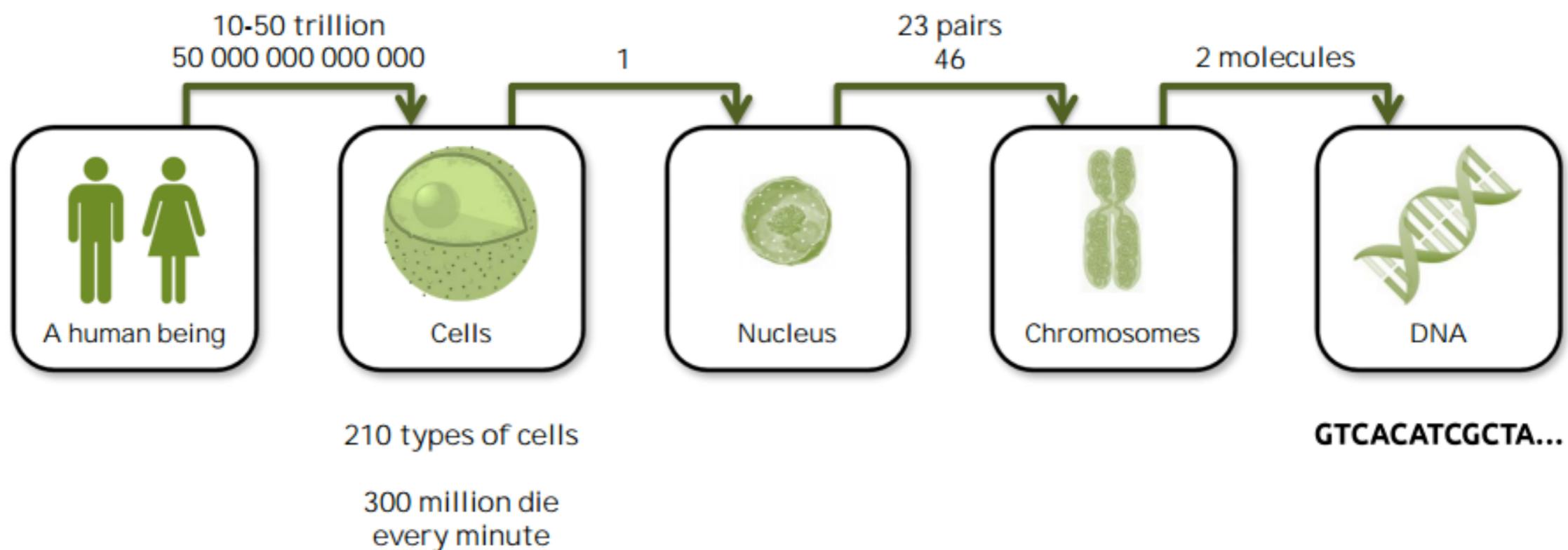


# Para isto usamos seu DNA!

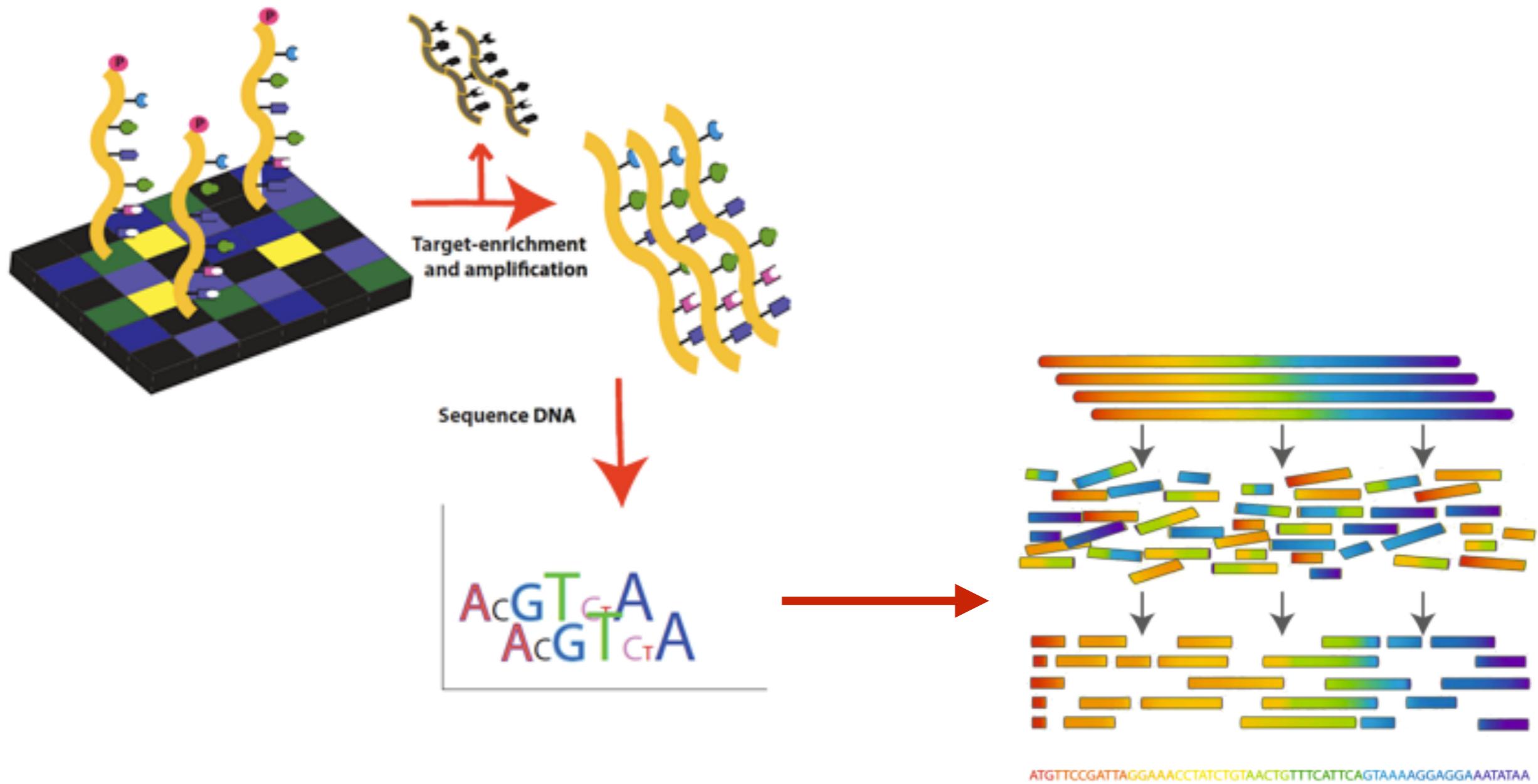




## From human beings to DNA



# Para isto usamos seu DNA!



# Para quem não entendeu lembra do Angelina Jolie effect ?

Startups | 31/08/2013 08:00

Comentários (0) Views (8.230)

+ Salvar notícia

## O efeito Angelina Jolie nas startups de mapeamento genético

O sequenciamento genético, que permitiu à atriz Angelina Jolie se livrar do risco de ter um câncer de mama, começa a se popularizar no Brasil graças a uma nova geração de startups

Guilherme Manechini , de **EXAME**

 Compartilhar 120

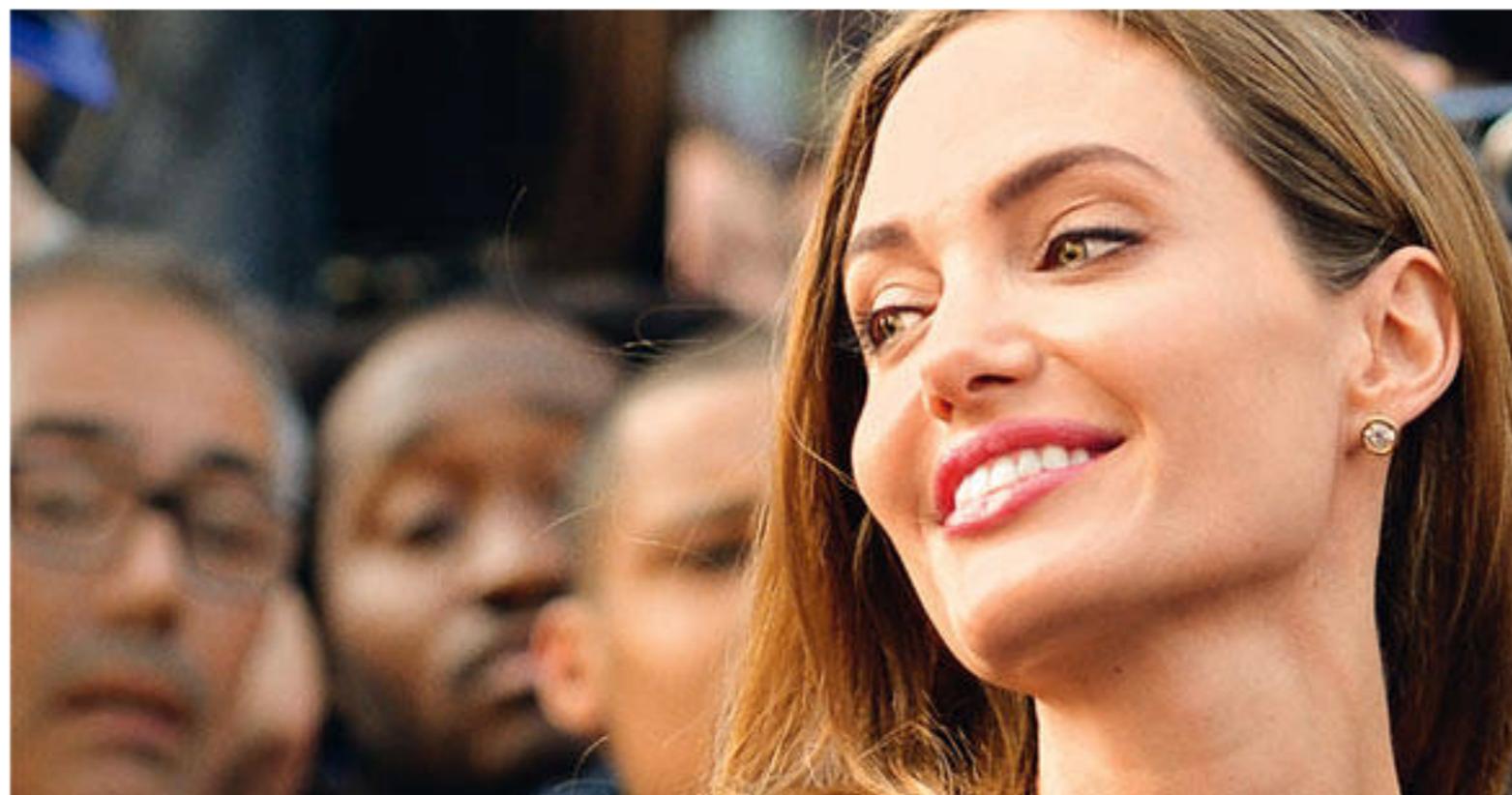
 Tweetar 6

 +1 0

 Share 5



Leon Neal/AFP Photo



# Para quem não entendeu lembra do Angelina Jolie effect ?

Startups | 31/08/2013 08:00

Comentários (0) Views (8.230)

+ Salvar notícia

O efeito A  
mapeamer

cial

Notícias

Entretenimento

Esportes

Vídeos

Rede Record

Serviços



O sequenciamento de mama, começa a

Guilherme Manechini , de

f Compartilhar 120

Você está aqui: Página Inicial/Notícias/Saúde

**Saúde**

Mais Médic

14/5/2013 às 17h49 (Atualizado em 15/5/2013 às 15h14)

## Exame feito em Angelina Jolie que detecta falha em gene custa em média R\$ 4.000

Revelação da atriz pode despertar interesse das mulheres no assunto, diz especialista

R7 Página inicial

f Recomendar 3

Tweetar 0

g+1 0

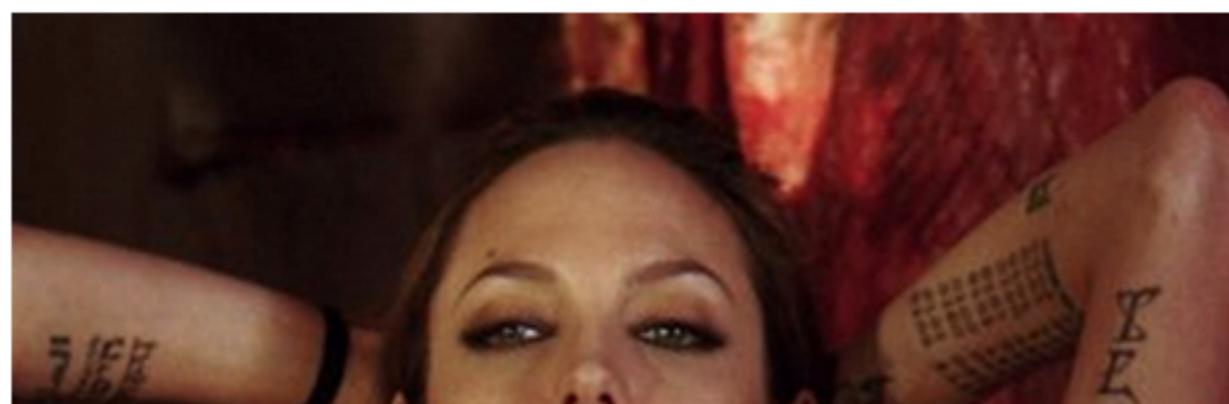
Pinit

RECEBA NOTÍCIAS NO SEU CELULAR

Texto: -A +A

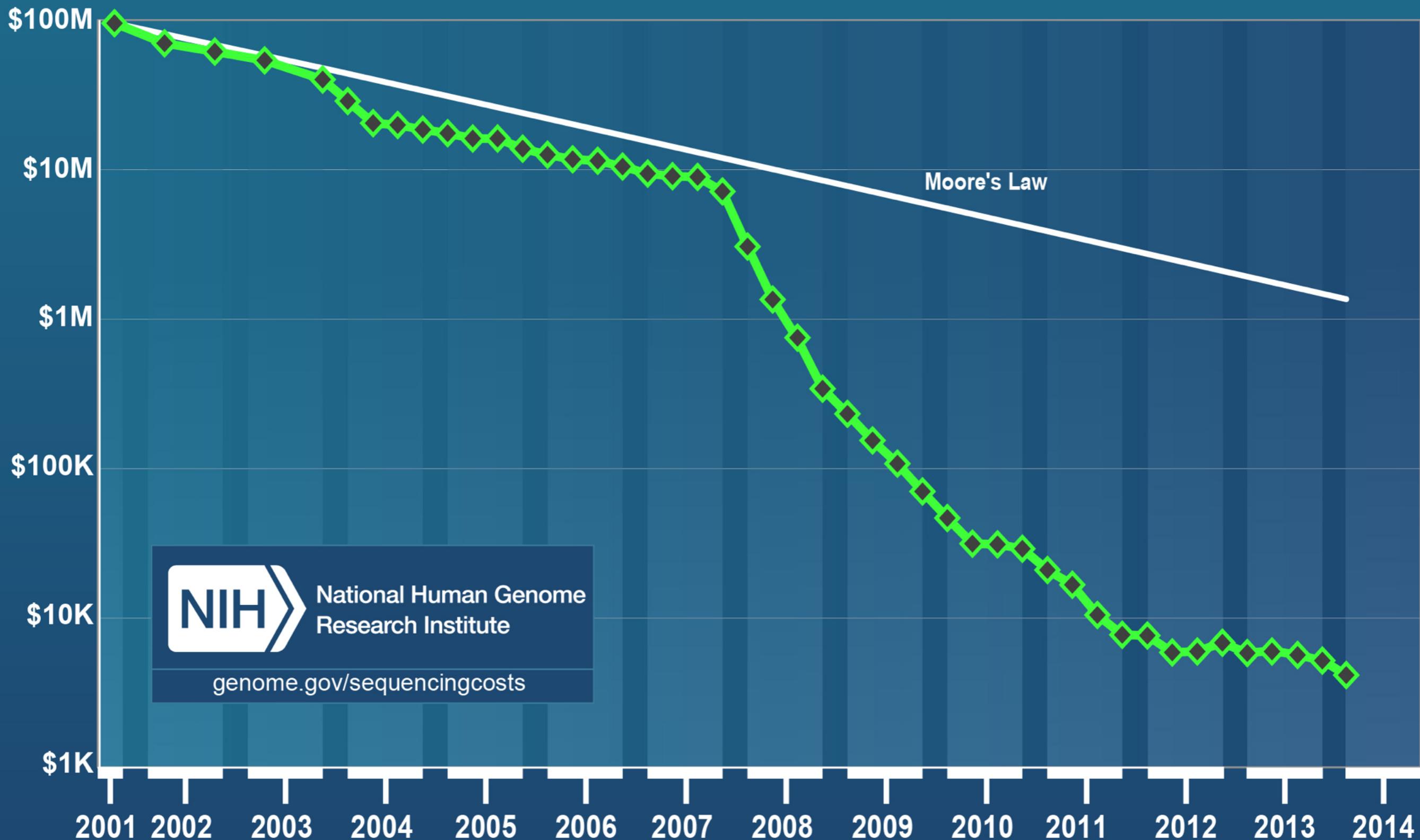


Do R7

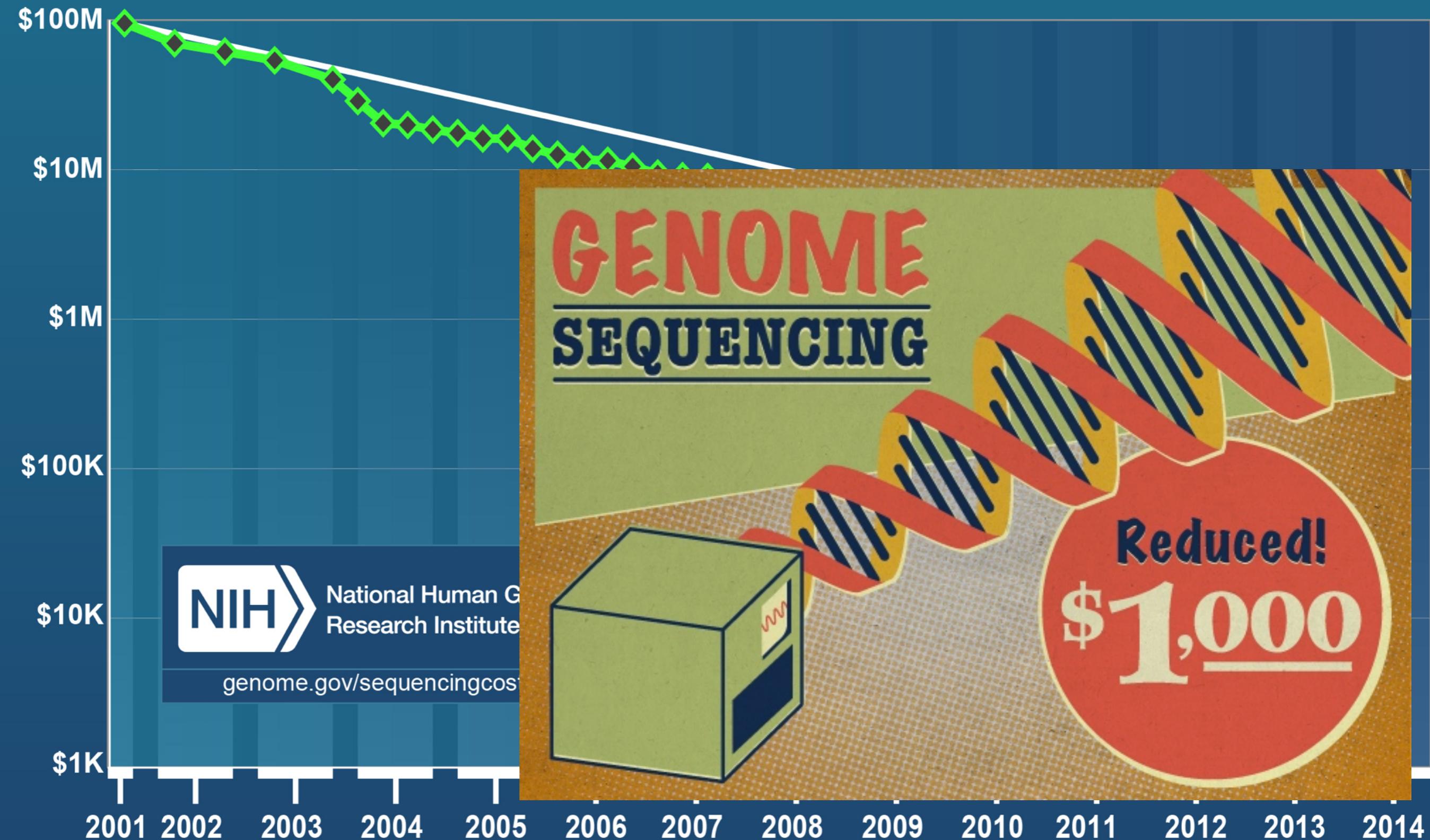


O exame que detecta a mutação genética realizado por Angelina Jolie custa em média R\$ 4.000 no Brasil. O exame de sangue permite saber se há alguma mutação cromossômica que predisponha ao câncer. Nesta terça-feira (14), a atriz revelou que se submeteu a uma cirurgia para a retirada dos dois seios, após

# *Cost per Genome*



## *Cost per Genome*



# NextGen Sequencing



MiSeq™

Welcome to Illumina MiSeq

SEQUENCE



illumina®

NCBI Reference Sequence: NM\_000374.3

## Homo sapiens uroporphyrinogen decarboxylase (UROD), mRNA

>gi|71051615|ref|NM\_000374.3| Homo sapiens uroporphyrinogen decarboxylase (UROD) , mRNA

```
ATGGCGACGCTTGGTCCCTACAGAAAGGGCGGAGCCTGGACTGGGGGCAGGCTCAGATTAGGTT  
AAATTGTGGATTGAGCTCGCAGTTACAGACAGCTGACCATGGAAGCGAATGGGTTGGACCTCAGGGTTT  
TCCGGAGCTGAAGAACATGACACATTCCCTGCGAGCAGCCTGGGGAGAGGAAACAGACTACACTCCCCTTGG  
TGCATGCCAGGCAGGCCGTTACTTACCAAGAGTTAGGGAAACCCGGGCTGCCAGGACTTTTCAGCA  
CGTTCGCTCTCCTGAGGCCTGCTGTGAACTGACTCTGCAGCCACTGCGTCGCTCCCTGGATGCTGC  
CATCATTTCCTCCGACATCCTGTTTACCCAGGCAGTGGCATGGAGGTGACCATGGTACCTGGCAAA  
GGACCCAGCTCCCAGAGCCATTAAGAGAACAGCAGGACCTAGAACGCCTACGGGATCCAGAAGTGGTAG  
CCTCTGAGCTAGGCTATGTGTTCCAAGCCATCACCCCTACCGACAACGACTGGCTGGACGTGTGCCGCT  
GATTGGCTTGCTGGTCCCCATGGACCCCTGATGACATACATGGTTGAGGGTGGCTCAAGCACCATG  
GCTCAGGCCAAGCGCTGGCTCTACAGAGACCTCAGGCTAGTCACCAGCTGCTTCGATCCTCACTGATG  
CTCTGGTCCCATATCTGGTAGGACAAGTGGTGGCTGGGCCAGGCATTGCAGCTGTTGAGTCCCATGC  
AGGGCATCTGGCCCACAGCTCTCAACAAAGTTGCACTGCCTTACATCCGTATGTGGCCAAGCAAGTG  
AAGGCCAGGTTGCCGGAGGCAGGCCTGGCACCAAGTGGCTGGGCCAGGCATTGCAGCTGTTGAGTCCCATGC  
CCCTGGAGGAGCTGGCCAAGCTGGCTATGAGGTGGTGGCTTGACTGGACAGTGGCCCCAAAGAAAGC  
CCGGGAGTGTGGGGAAAGACGGTGACATTGCAGGGCAACCTGGACCCCTGTGCCTGTATGCATCTGAG  
GAGGAGATGGGCAGTTGGTAGCAGATGCTGGATGACTTGGACCACATCGCTACATTGCCAACCTGG  
GCCATGGCTTATCCTGACATGGACCCAGAACATGTGGCGCCTTGTGGATGCTGTGCATAAACACTC  
ACGTCTGCTTCGACAGAACTGAGTGTATACCTTACCCCTCAAGTACCAACTAACACAGATGATTGATCGTT  
TCCAGGACAATAAAAGTTGGAGTTGAACATTGTGTAGTTGTGAAAGATTGTGCCCATATCC  
TCAGTTCTTCTAGCCTCTGCTCCTCCCTGGGAACCCCTCTATATCCTT
```

# Tamanho

Sequenced  
Genome

180 GB

“Raw” data  
FASTQ format

```
> HG00339 read123 300bp
ATGCATCGTACGCTCATCATGCAGC
+
!''*(((***+))%%%++)(%%%).1 ***
> HG00339 read456 300bp
TCGTACGCTCATCATGCAGCAGGTC
+
%%%).1 *****))**5CCF>>>>CC
...
> HG00339 read999 300bp
CAGTACGTCAGTGCTAGCTAGCTAC
+
)(%%%).1 *****))**5CCCC?CCC65
```



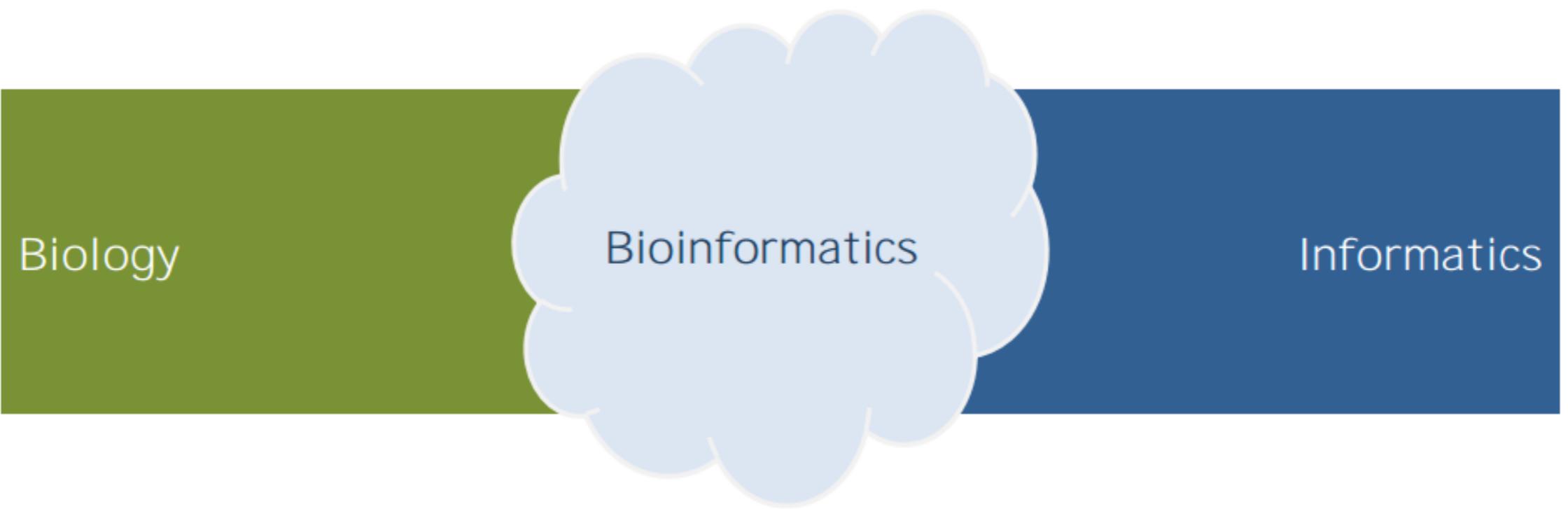
# Volume

How much disk space do we need to store all genomes of an entire population?

	Population	Compressed	Size with 2-bits	Size with 8-bits	WGS
1-person	1	7 MB	770 MB	3 GB	180 GB
Santa Maria, RS	270 K	1.8 TB	200 TB	790 TB	46 PB
Porto Alegre, RS	1.5 M	10 TB	1.1 PB	4.3 PB	260 PB
Rio Grande do Sul	10.7 M	72 TB	7.7 PB	30 PB	1.8 EB
São Paulo, SP	11.3 M	75 TB	8.1 PB	32 PB	1.9 EB
Brazil	201 M	1.3 PB	144 PB	575 PB	34 EB
Latin America	589 M	3.8 PB	422 PB	1.6 EB	100 EB
World	7 B	45 PB	5 EB	20 EB	1.1 ZB



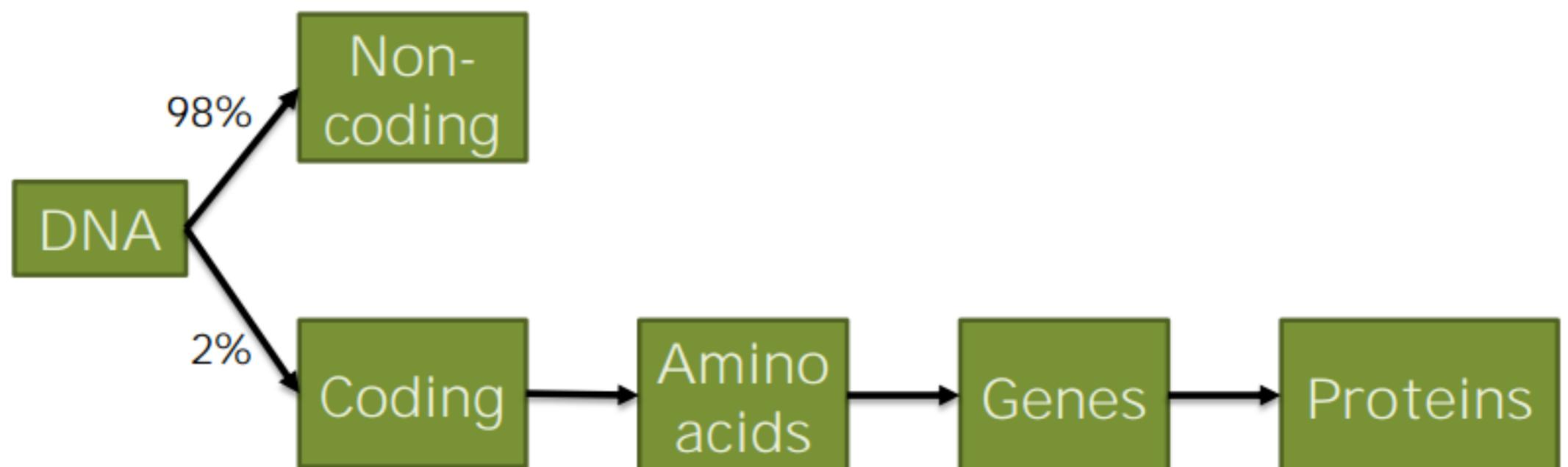
# What is bioinformatics ?



Bioinformatics is an informatics discipline for storing, retrieving, organizing and analyzing biological data.



# Meaning





# Meaning

CATTCATCATGCATACGGACTGCAGCAGGT

CATTCATTATGCATACGGACTGCAGCAGGT



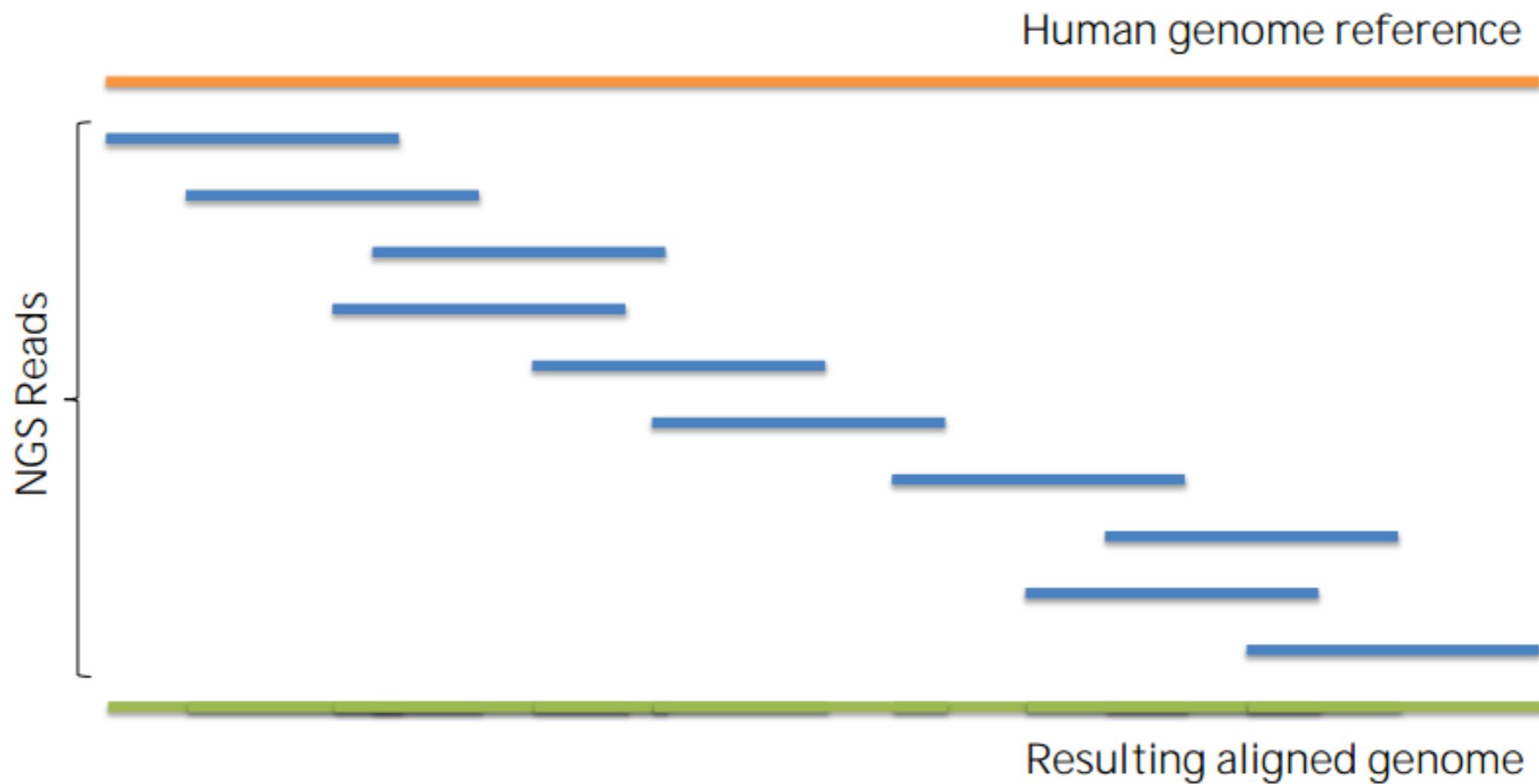
# Meaning

CATTCAT**C**ATGCATAACGGACTGCAGCAGGT

CATTCATT**T**ATGCATAACGGACTGCAGCAGGT



# Meaning



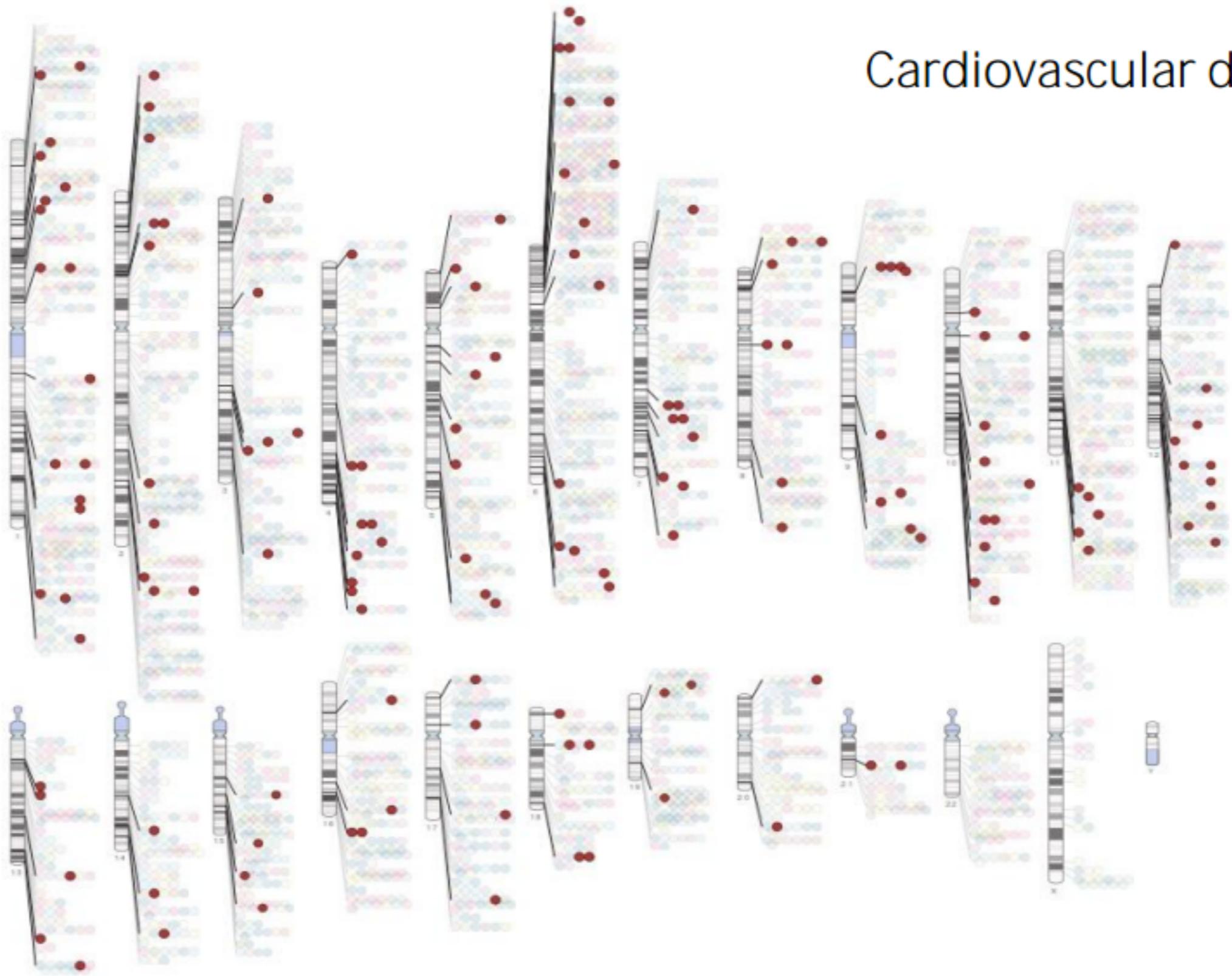


# Meaning





# Meaning



# Variant Analysis

Assertion and evidence details D

Go to: [▼](#) [▲](#)

Clinical Assertions Evidence Help

**Germline**

Clinical significance (Last evaluated)	Review status (Assertion method)	Collection method	Condition(s) (Mode of inheritance)	Origin	Citations	Submitter (Last submitted)	Submission accession
Pathogenic (Mar 3, 2004)	reviewed by professional society (evidence-based review)	literature only	Cystic fibrosis (Autosomal recessive inheritance) <a href="#">[MedGen]</a> <a href="#">[Orphanet]</a> <a href="#">[OMIM]</a>	germline		<a href="#">American College of Medical Genetics and Genomics (ACMG)</a> (Jun 3, 2013)	SCV000071408
Pathogenic (Mar 28, 2013)	reviewed by expert panel (evidence-based review)	literature only	Cystic fibrosis <a href="#">[MedGen]</a> <a href="#">[Orphanet]</a> <a href="#">[OMIM]</a>	germline	<a href="#">PubMed (1)</a>	<a href="#">CFTR2</a> (May 29, 2013)	SCV000071513
Pathogenic (Feb 13, 2013)	classified by single submitter (literature only)	literature only	Cystic fibrosis <a href="#">[MedGen]</a> <a href="#">[Orphanet]</a> <a href="#">[OMIM]</a>	germline	<a href="#">PubMed (2)</a>	<a href="#">OMIM</a> (Dec 30, 2010)	SCV000027757
not provided (Feb 1, 2013)	not classified by submitter (literature only)	literature only	Cystic fibrosis <a href="#">[MedGen]</a> <a href="#">[Orphanet]</a> <a href="#">[OMIM]</a>	germline	<a href="#">PubMed (2)</a>	<a href="#">Invitae</a> (Mar 30, 2013)	SCV000075041

Last Updated: Jun 20, 2014

# Variant Analysis

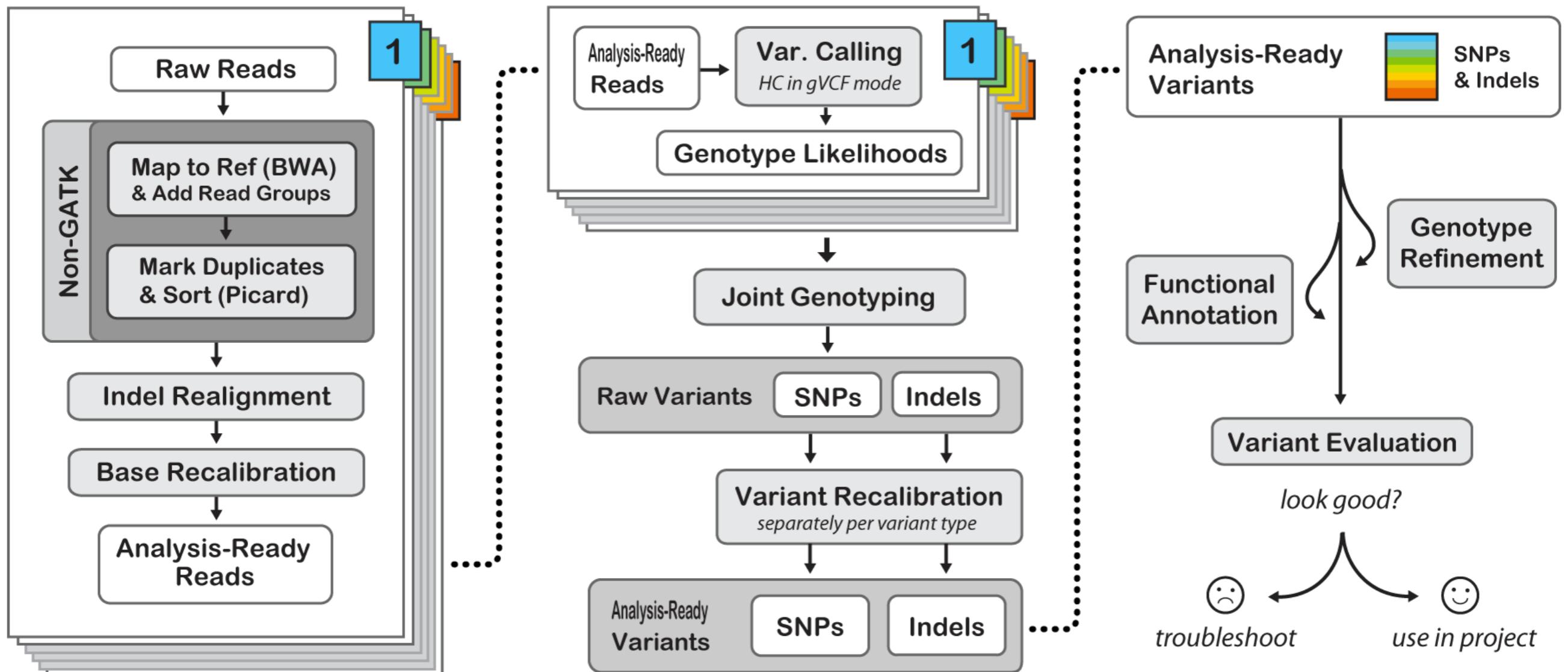
Data Pre-processing

>>

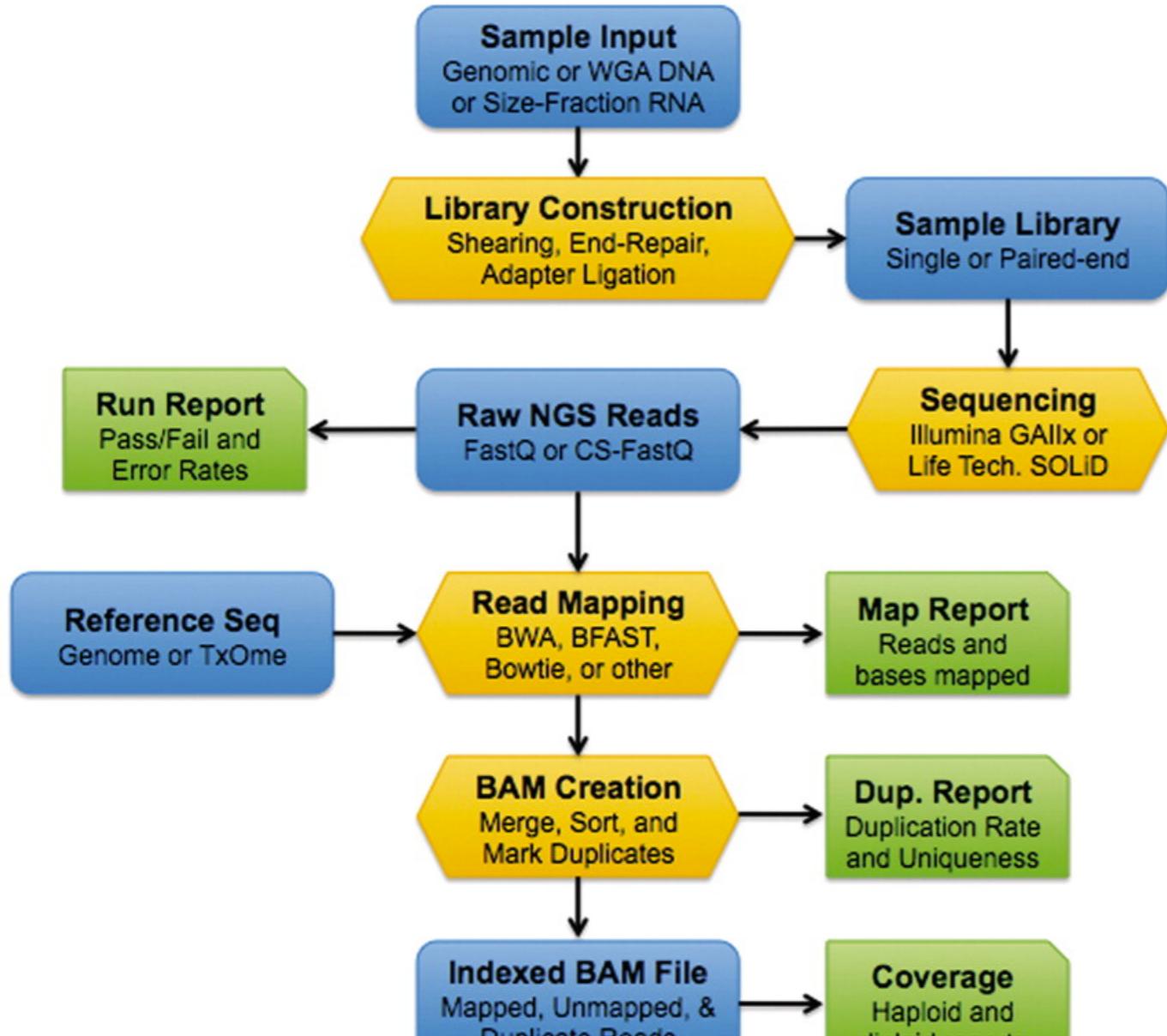
Variant Discovery

>>

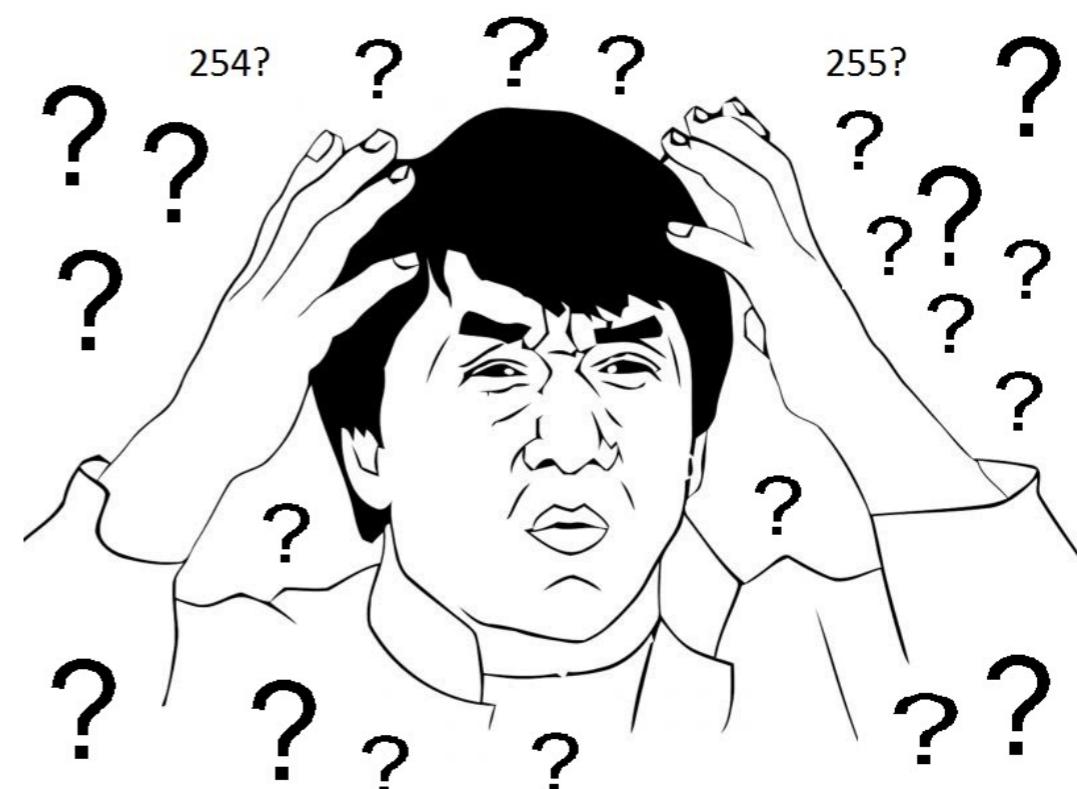
Preliminary Analyses



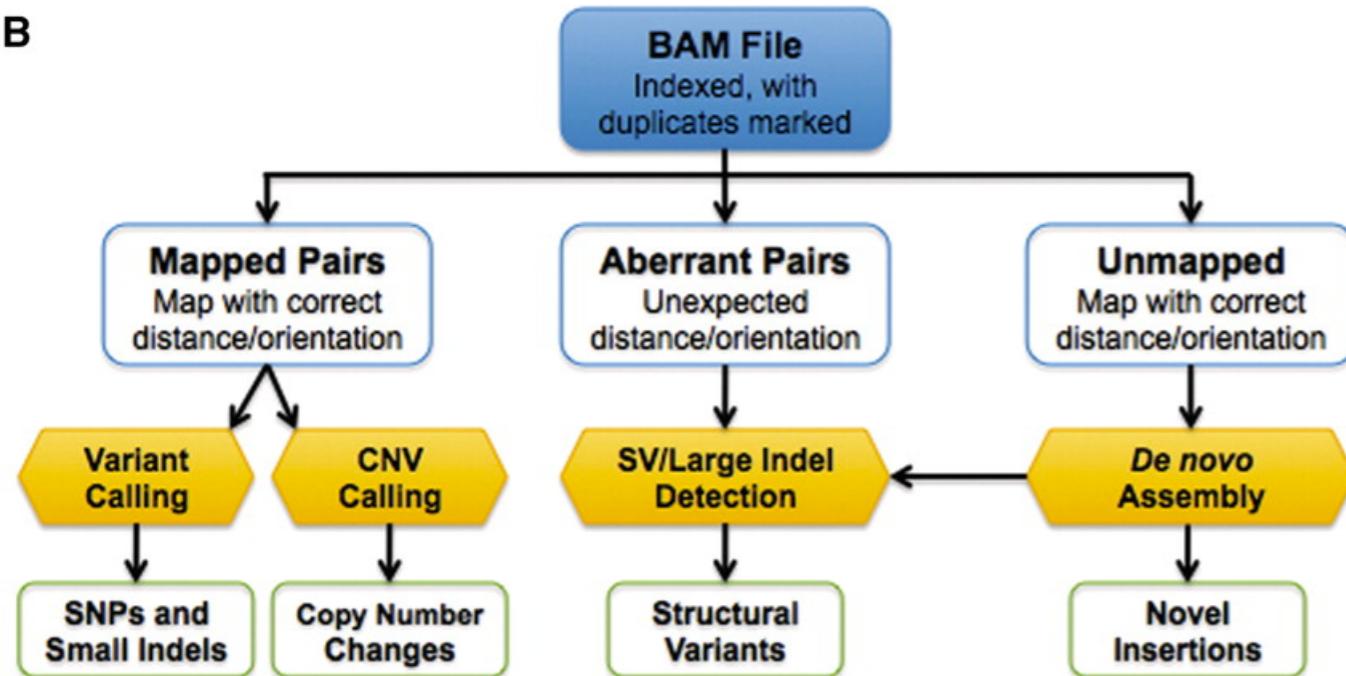
A



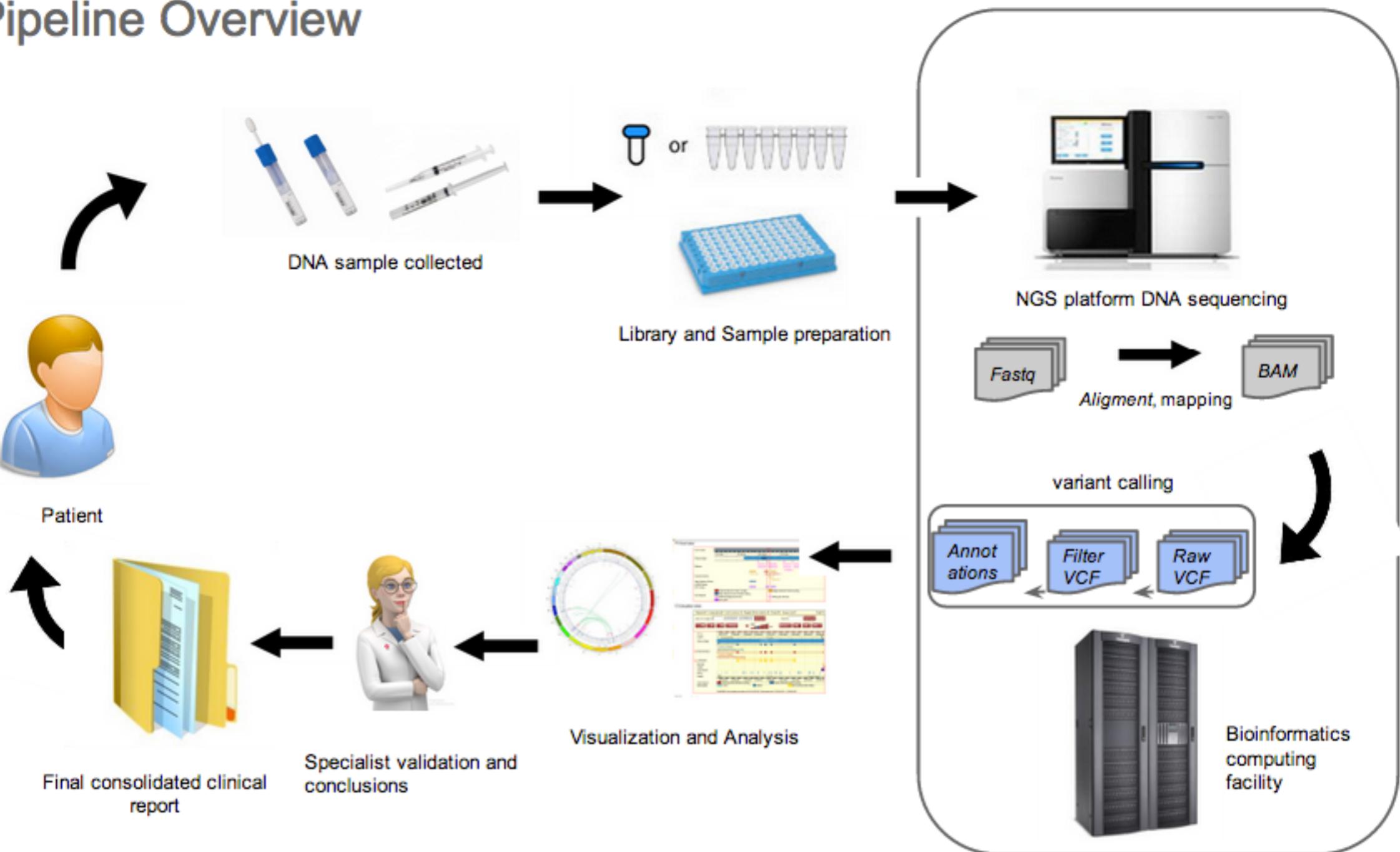
is



B



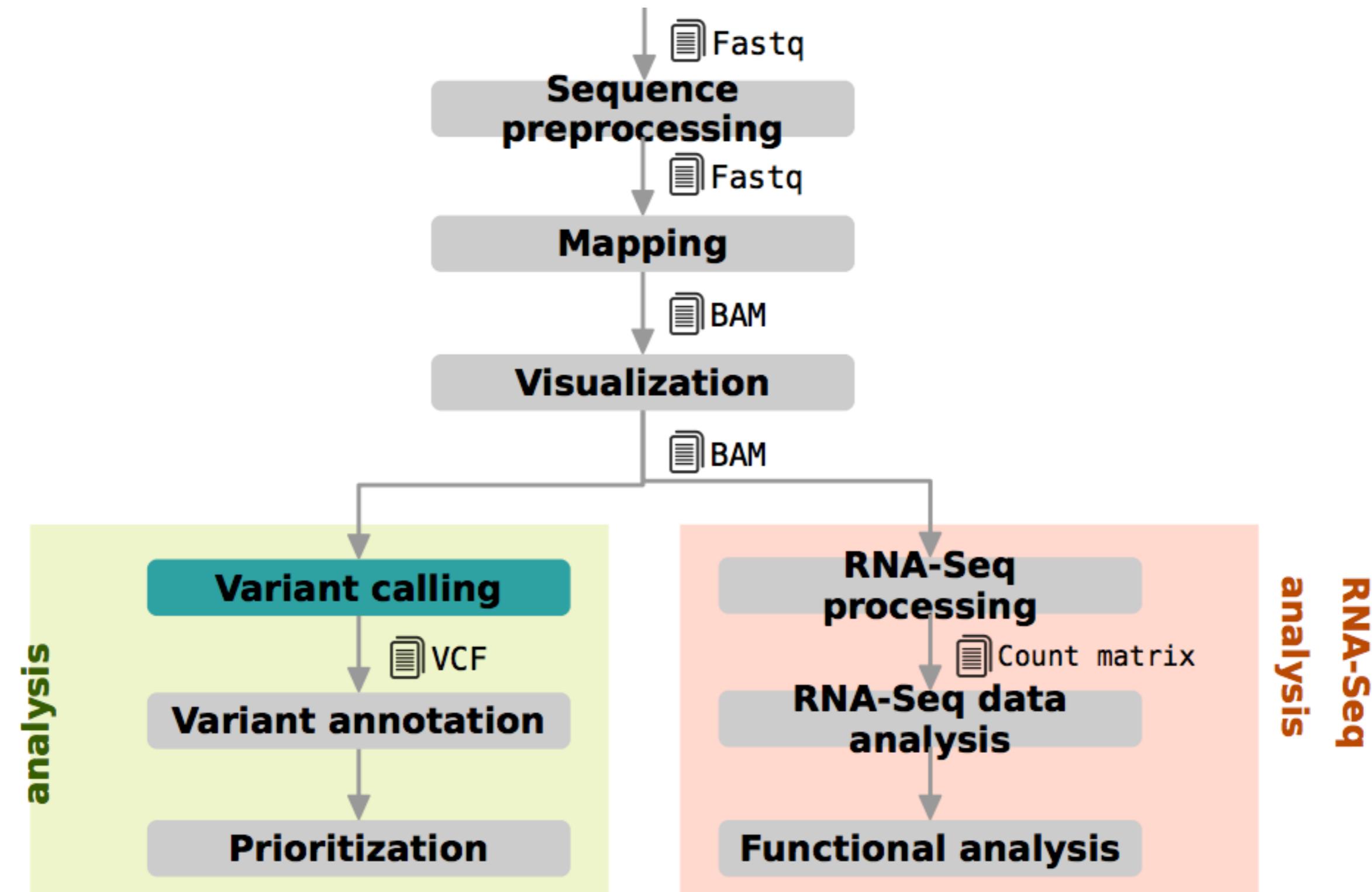
# Pipeline Overview



# Agenda

- › **Dia 1 (26/02)** : Introdução a Tecnologias NGS e Hands-on de Linux/Unix e Shell.
- › **Dia 2 (28/02)**: Controle de Qualidade; Alinhamento dos dados; visualização dos dados NGS ;
- › **Dia 3 (02/02 e 04/02)**: Análise de Cobertura e Chamada de Variantes
- › **Dia 4 (04/02)**: Priorização de variantes, análises, big data e próximos passos.

# Overview



# NGS, Big Data.



# Volume, por exemplo

How much disk space do we need to store all genomes of an entire population?

	Population	Compressed	Size with 2-bits	Size with 8-bits	WGS
1-person	1	7 MB	770 MB	3 GB	180 GB
Santa Maria, RS	270 K	1.8 TB	200 TB	790 TB	46 PB
Porto Alegre, RS	1.5 M	10 TB	1.1 PB	4.3 PB	260 PB
Rio Grande do Sul	10.7 M	72 TB	7.7 PB	30 PB	1.8 EB
São Paulo, SP	11.3 M	75 TB	8.1 PB	32 PB	1.9 EB
Brazil	201 M	1.3 PB	144 PB	575 PB	34 EB
Latin America	589 M	3.8 PB	422 PB	1.6 EB	100 EB
World	7 B	45 PB	5 EB	20 EB	1.1 ZB

# Hands-On (checklist)

## › **Acesso ao servidor**

› Todos os softwares já foram instalados, portanto não será necessário baixá-los. Mas é importante aprender os conceitos básicos de Linux, pois maior parte dos comandos para instalação e execução são neste ambiente.

## › **Entrega da Planilha de Atalhos**

## › **Apresentação dos alunos**



---

**Dando boas vindas!**

Marcel Caraciolo, CTO  
[marcel@genomika.com.br](mailto:marcel@genomika.com.br)