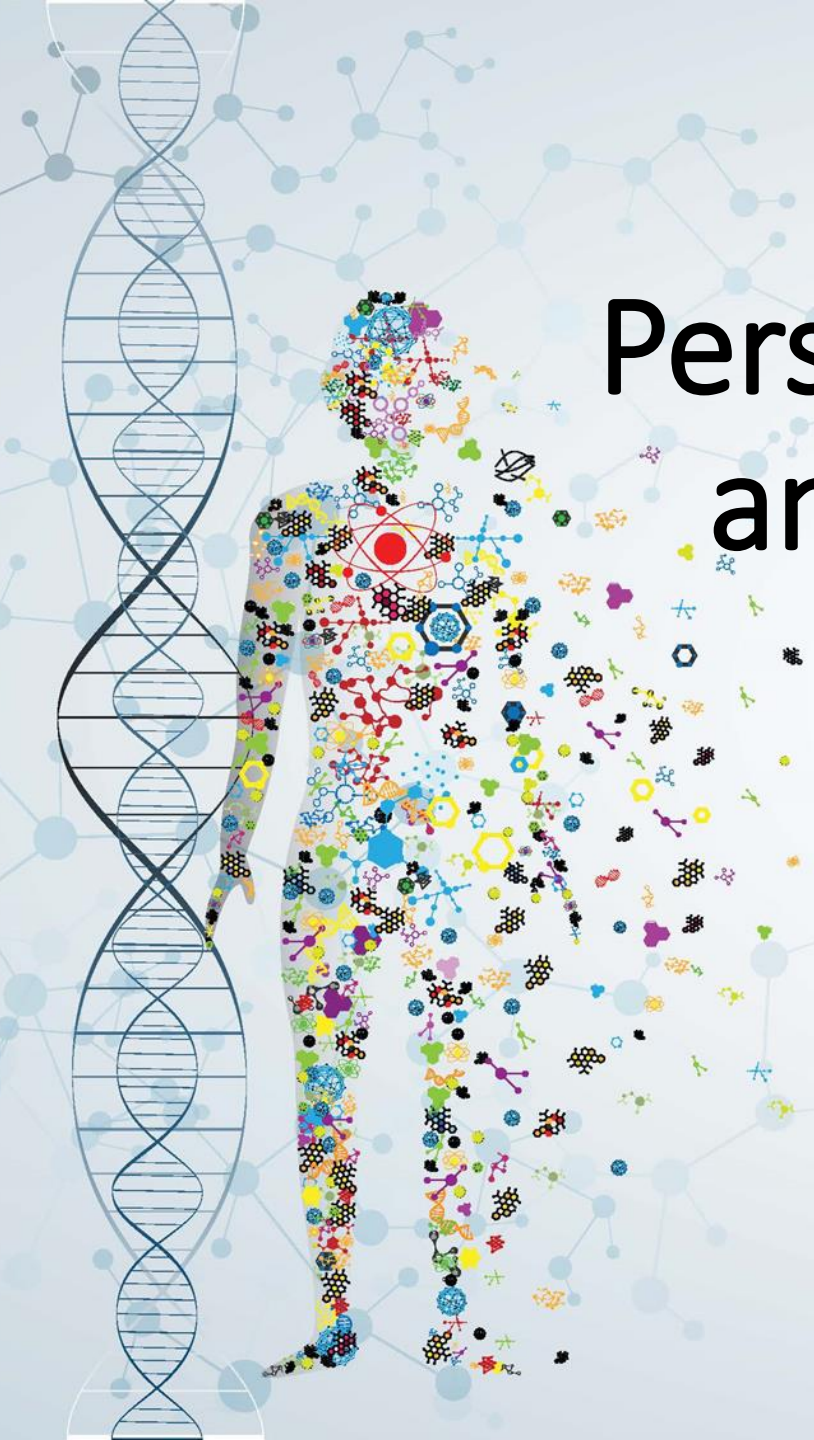


Personalized Medicine and Genomic Data Science

Ezgi Karaesmen

PhD candidate at OSU, College of Pharmacy

What is personalized medicine?

- Also termed **precision medicine** (oh yes, scientists always manage to come up with several terms for one thing)
- According to the NIH's Precision Medicine Initiative, it is:

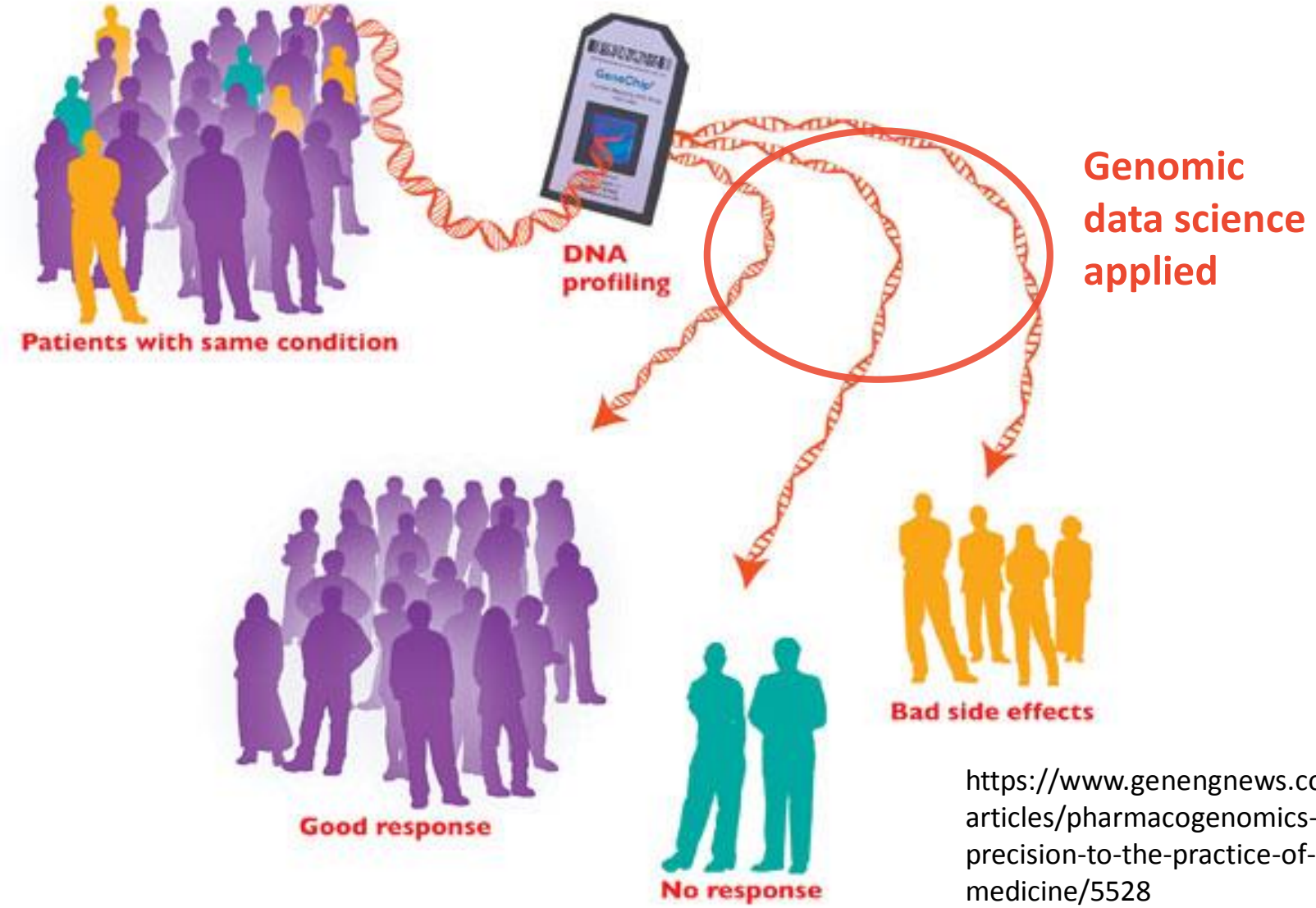
"An emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person."

What is pharmacogenomics?

- Pharmacogenomics is a part of precision medicine.
- According to NIH:

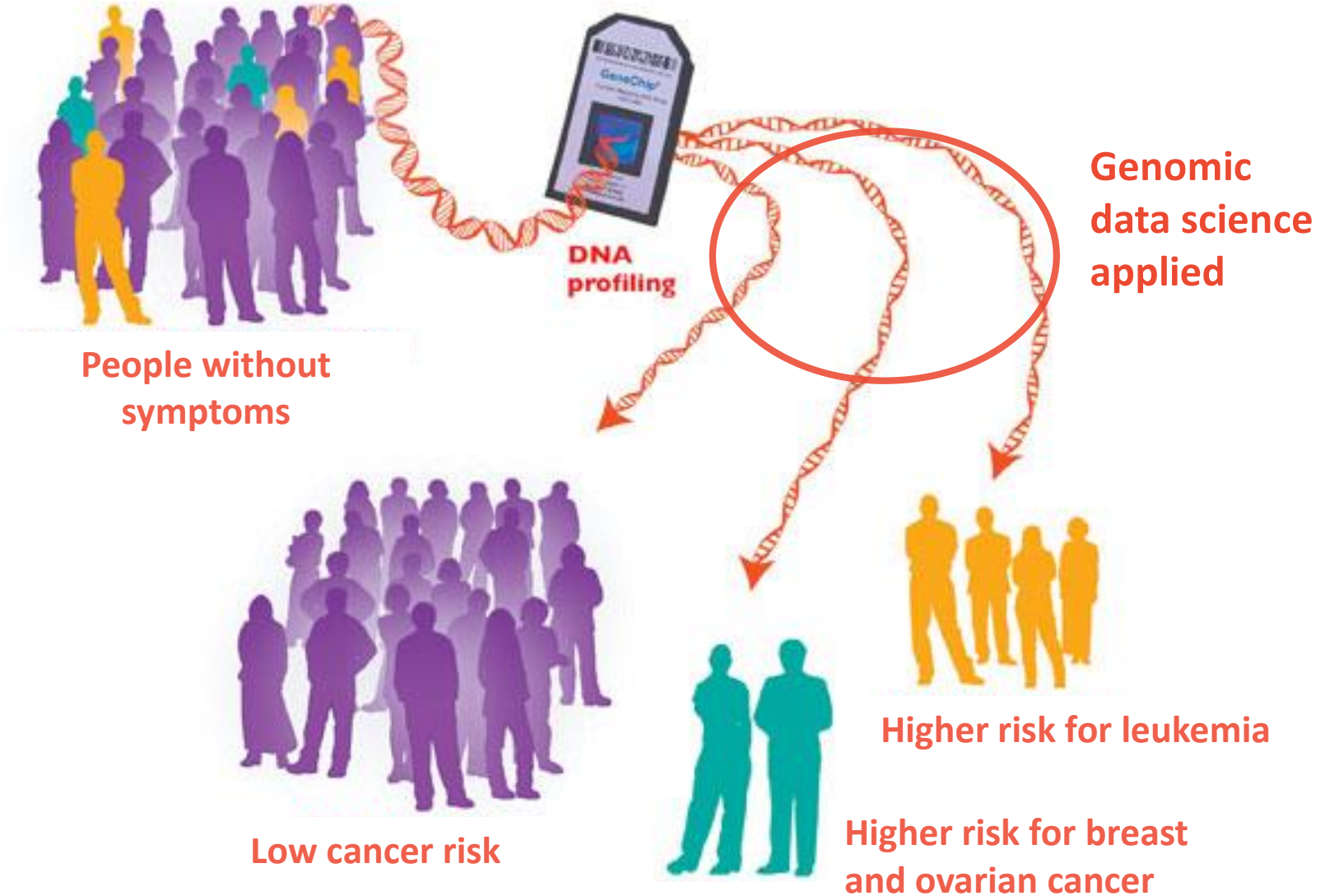
  “Pharmacogenomics is the study of how genes affect a person’s response to drugs. This relatively new field combines pharmacology (the science of drugs) and genomics (the study of genes and their functions) to develop effective, safe medications and doses that will be tailored to a person’s genetic makeup.”

What is pharmacogenomics?

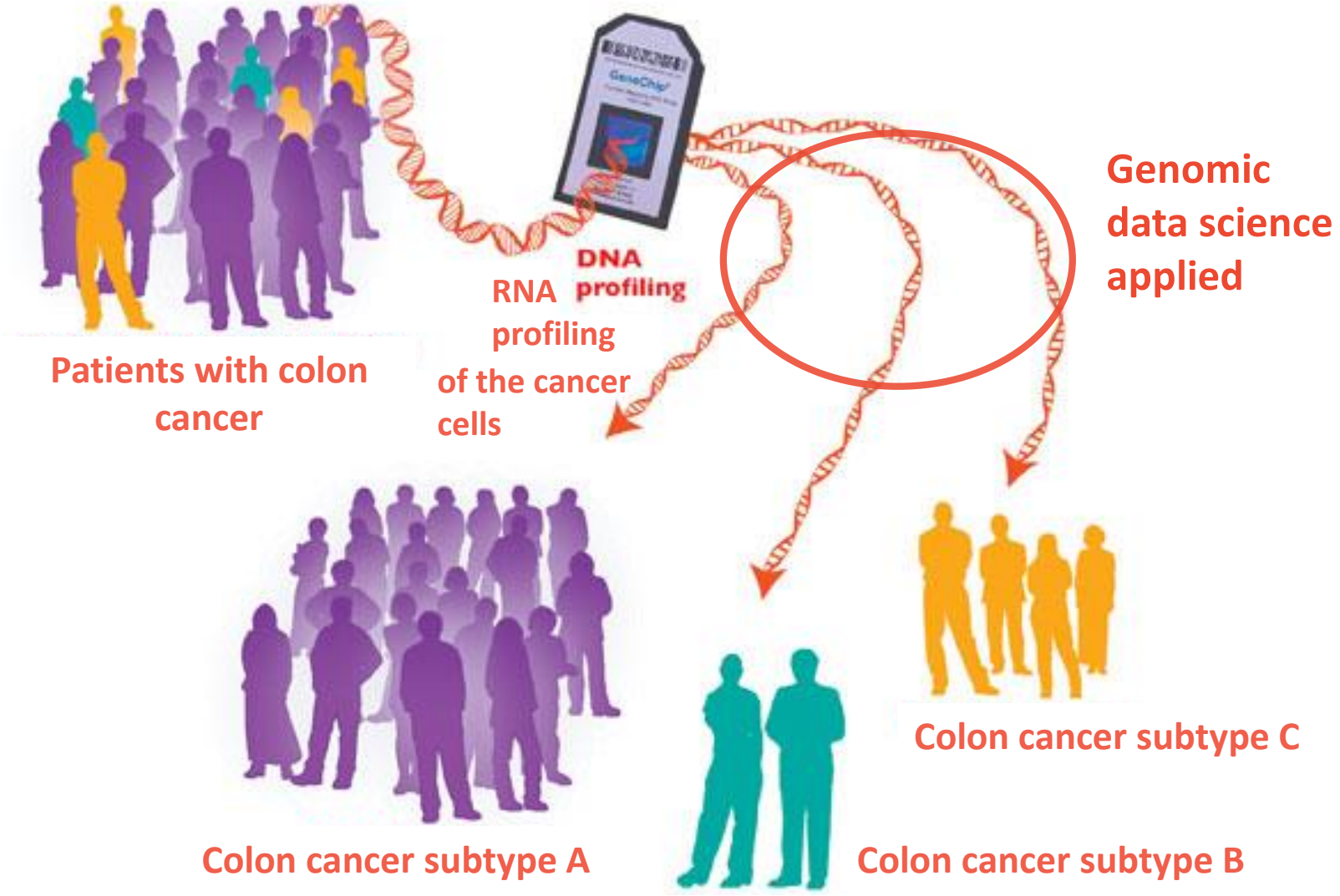


https://www.genengnews.com/gen-articles/pharmacogenomics-adds-precision-to-the-practice-of-medicine/5528

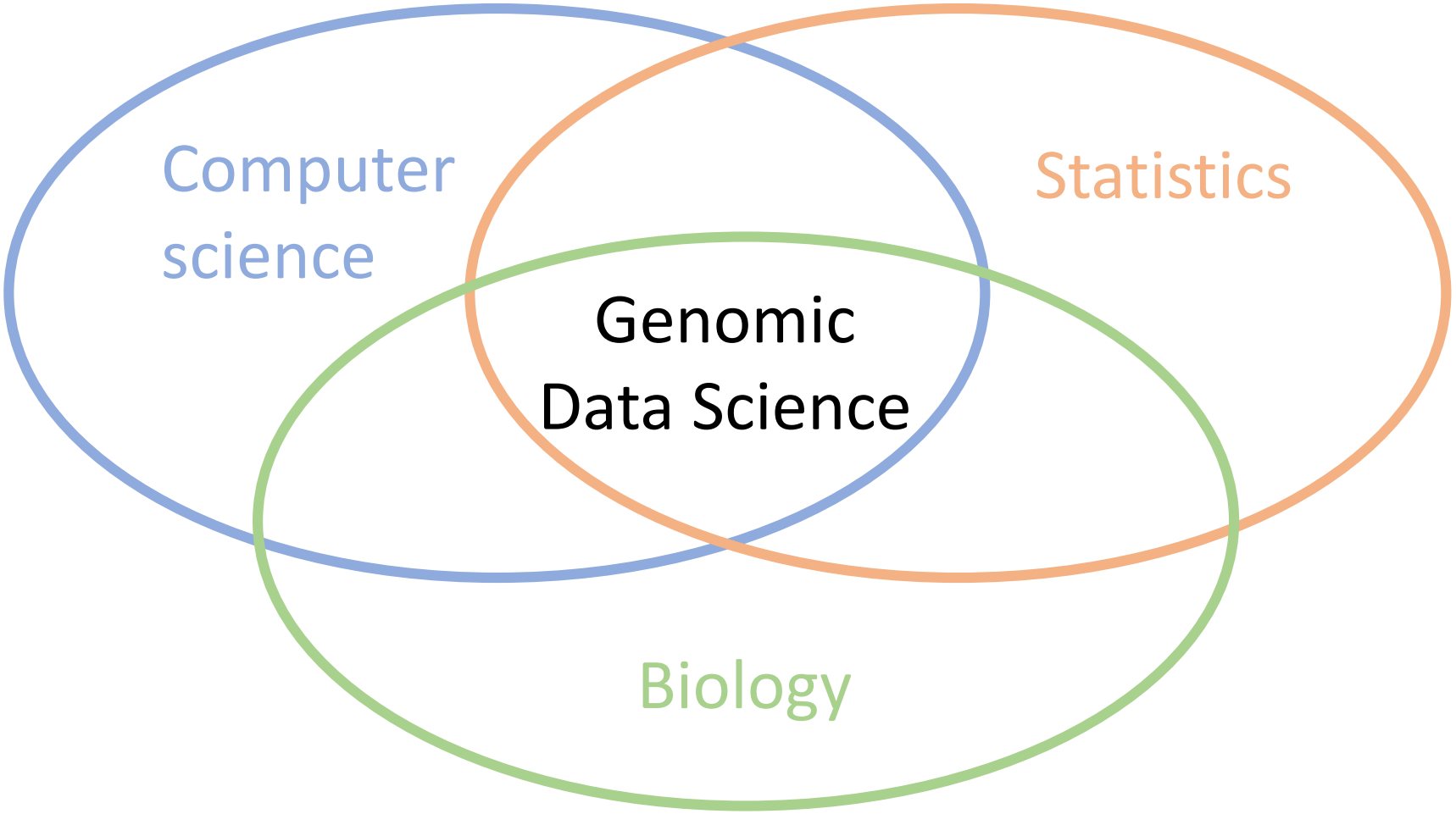
Precision medicine isn't just pharmacogenomics!



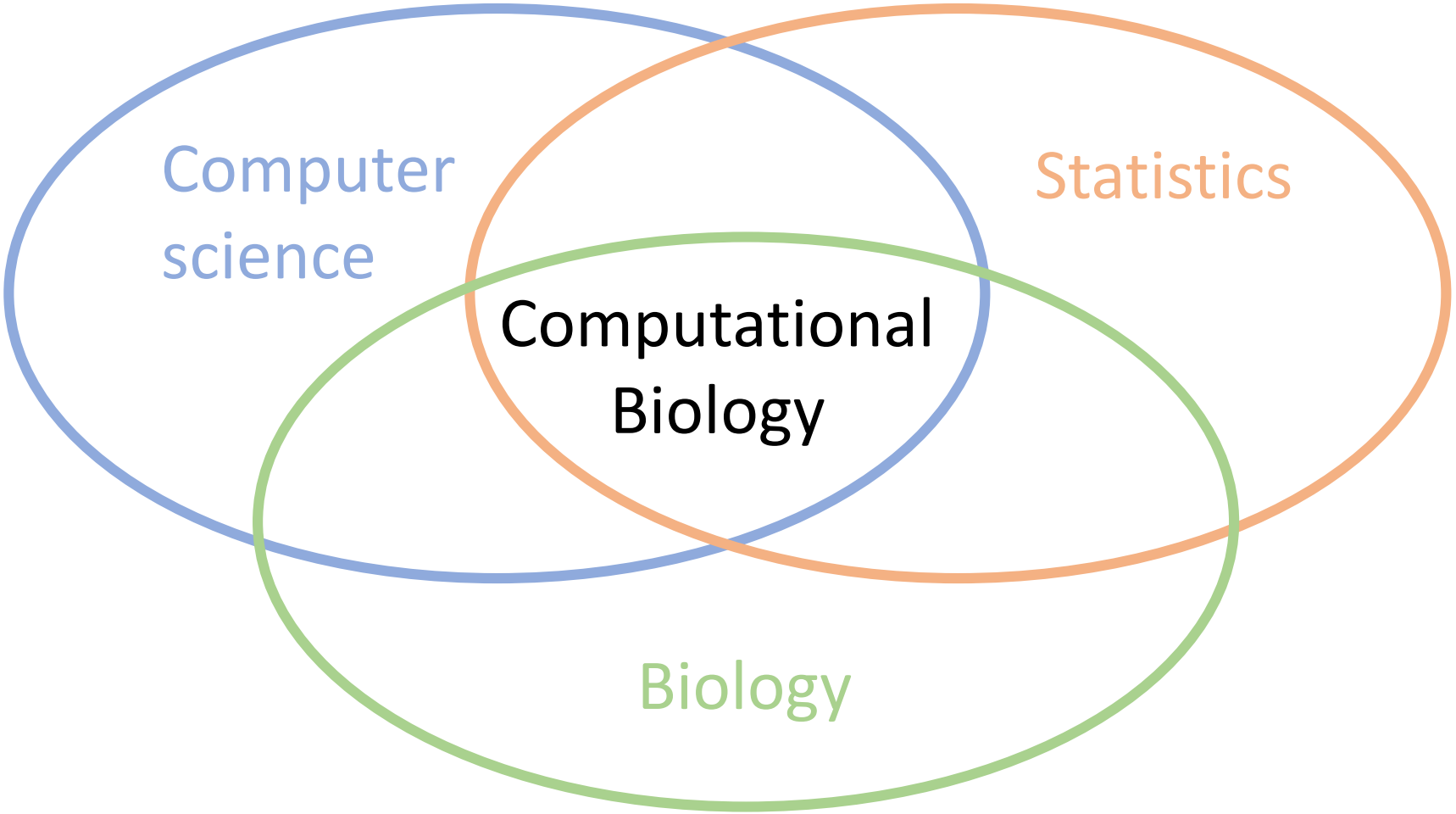
Precision medicine isn't just pharmacogenomics!



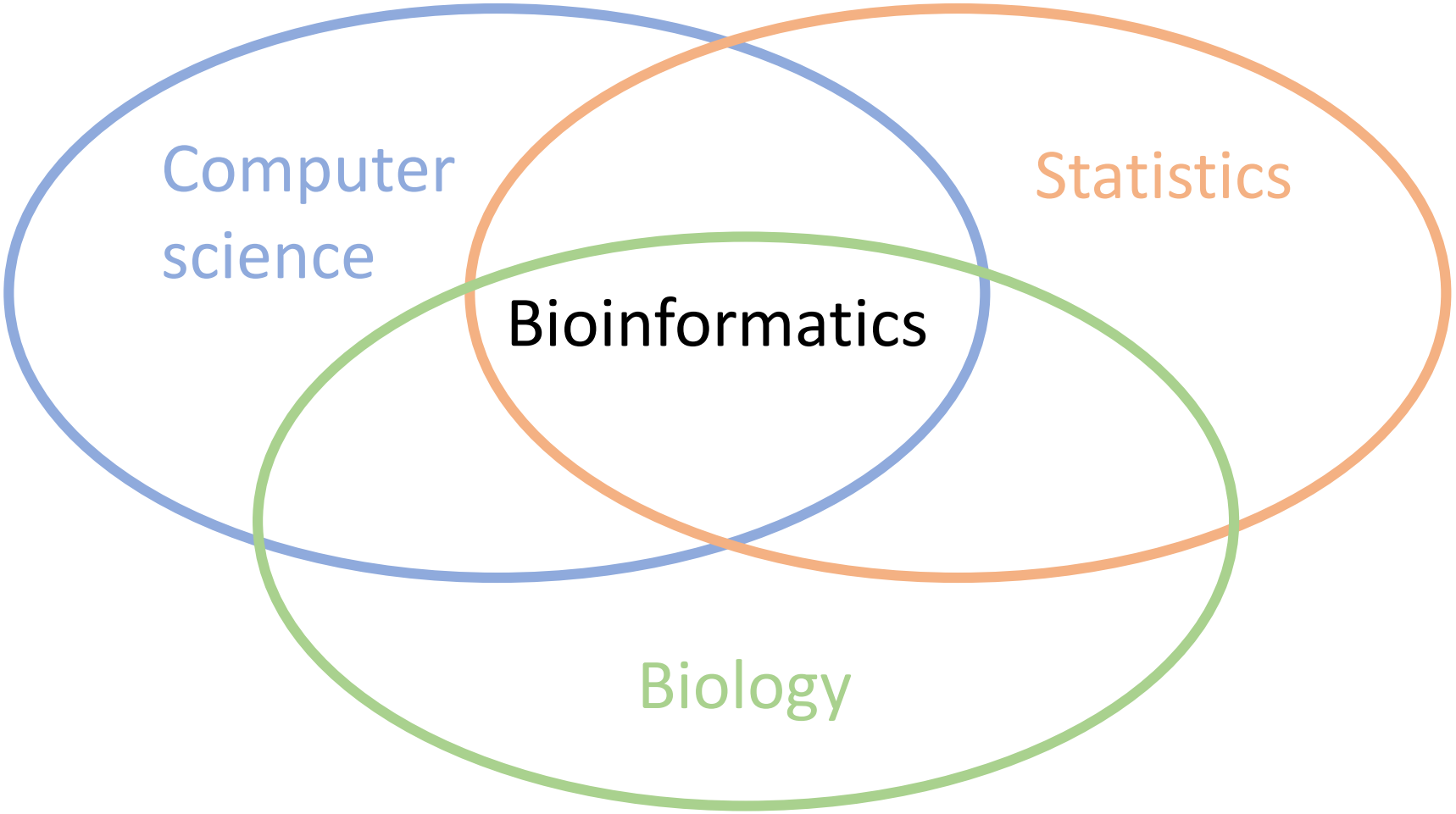
What is genomic data science?



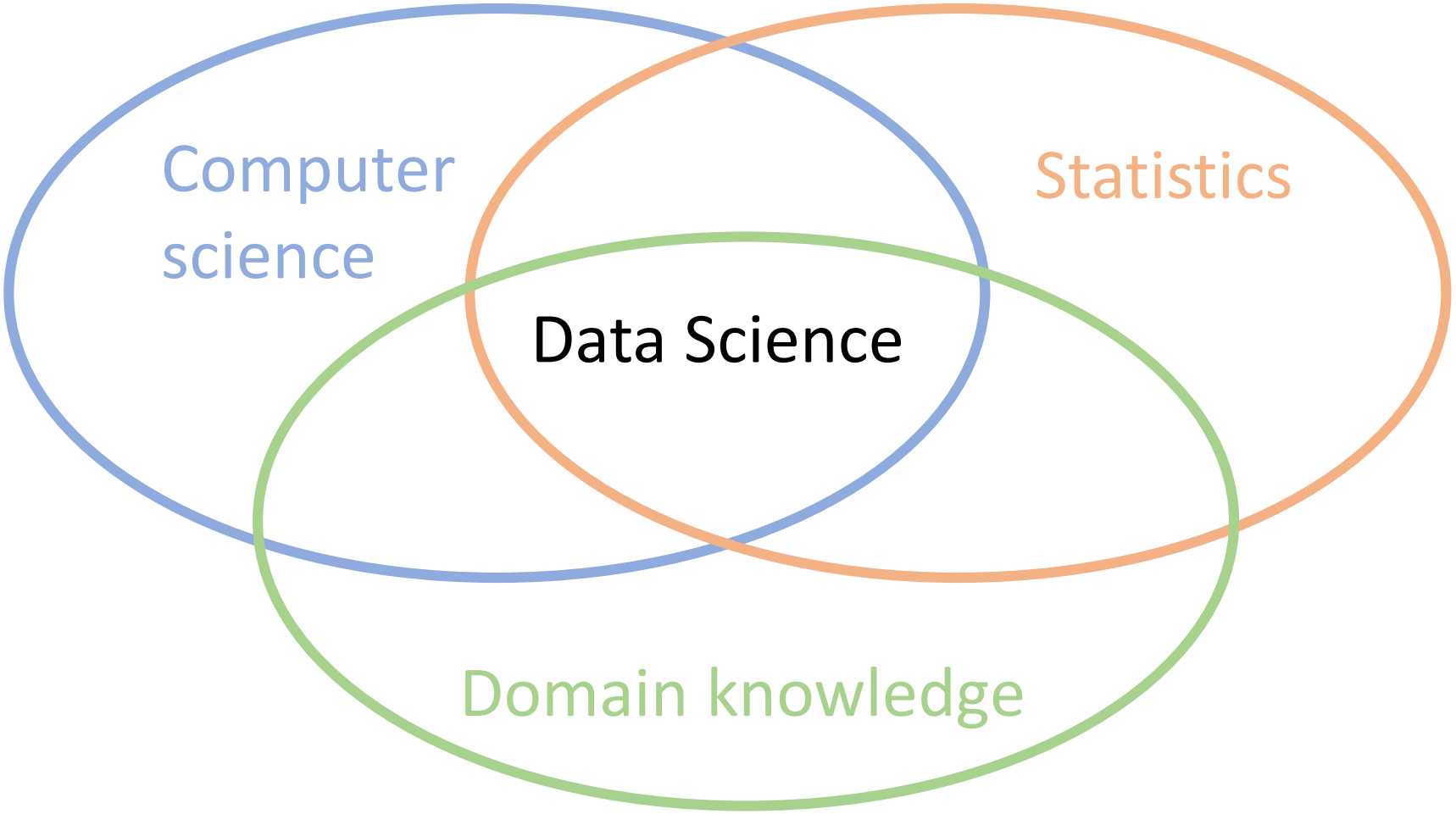
What is genomic data science?



What is genomic data science?

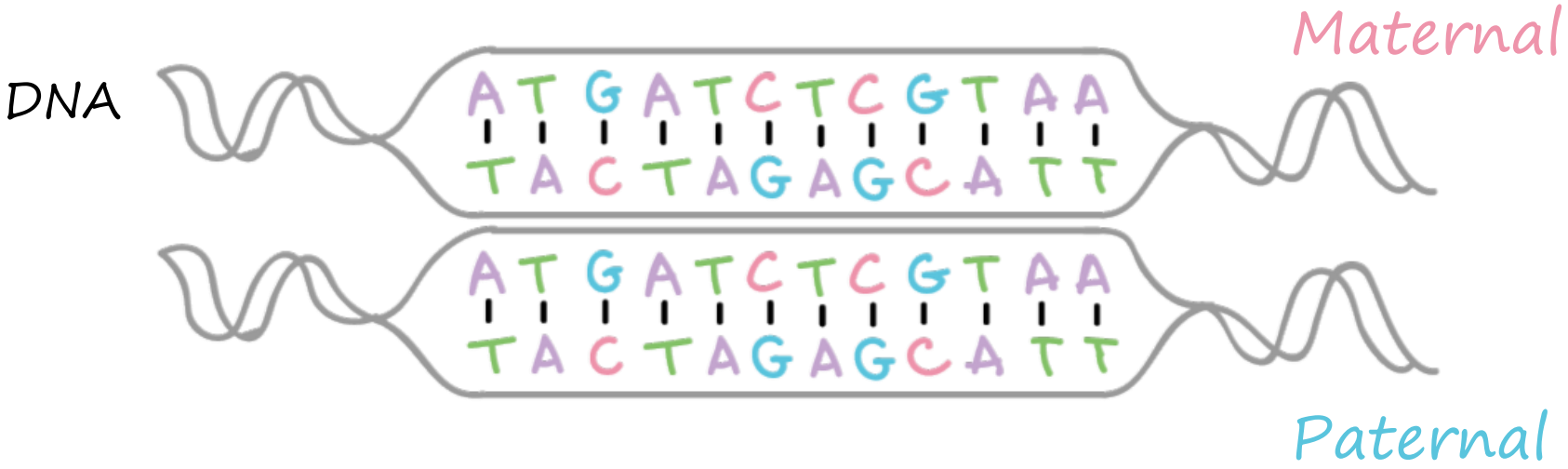


What is genomic data science?

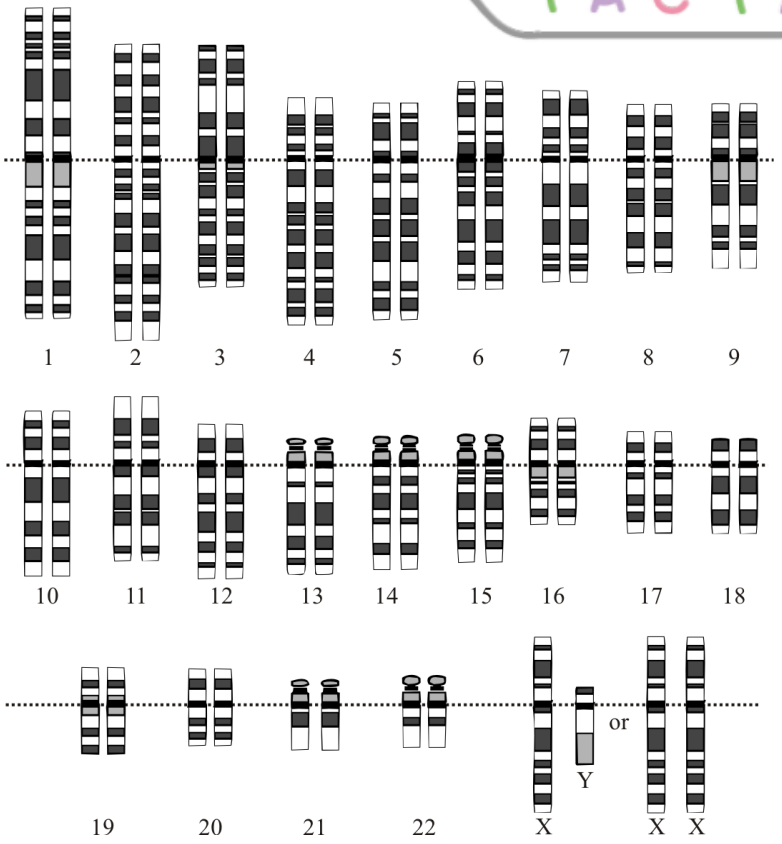


High school biology refresher

High school biology refresher

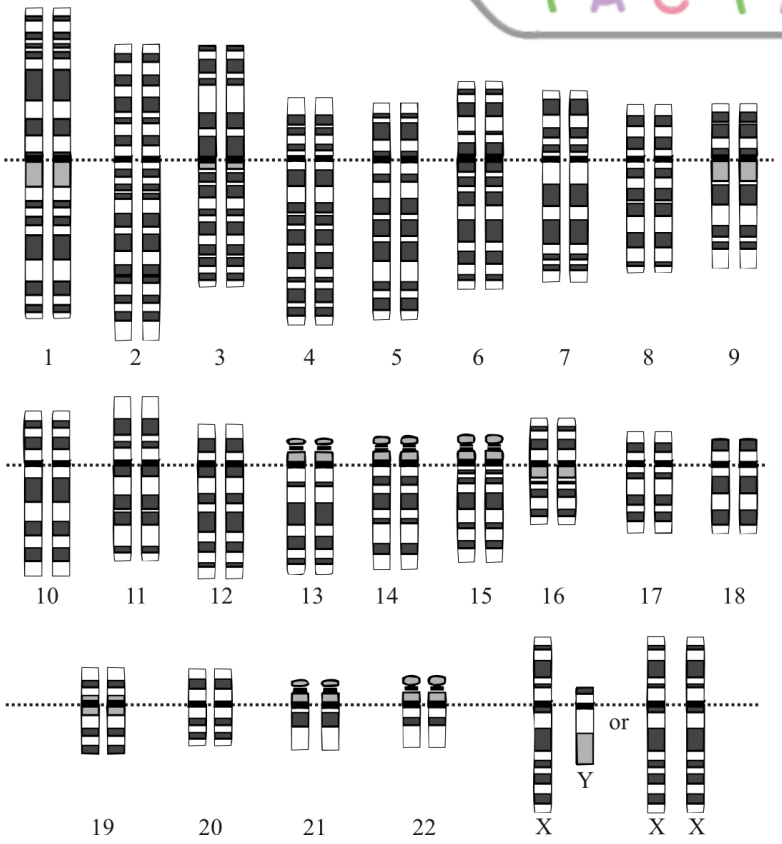


High school biology refresher



National Institute of Health and National Human Genome Research Institute

High school biology refresher



Size of the human genome

Bases

3 billion × 2 = 6 billion

Chromosomes

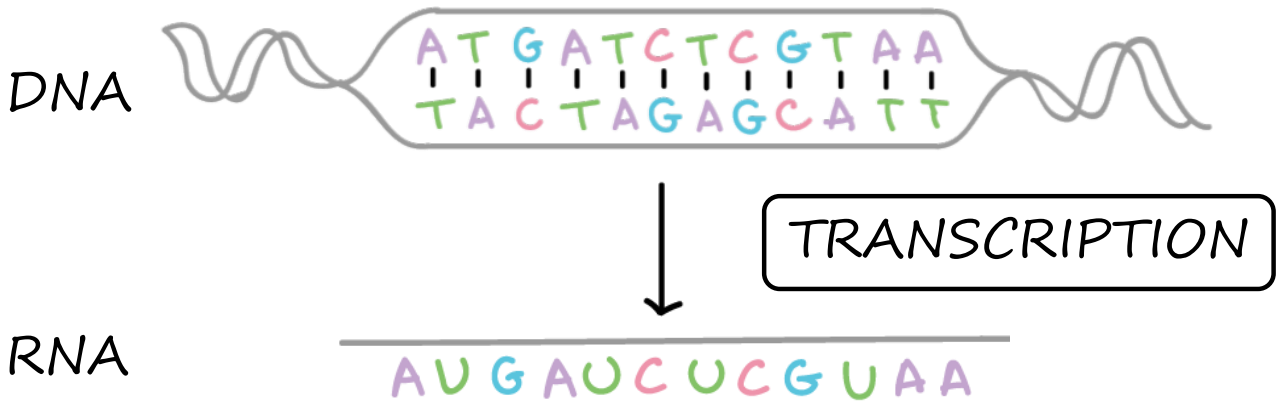
22 × 2 + (XX or XY)

National Institute of Health and National Human Genome Research Institute

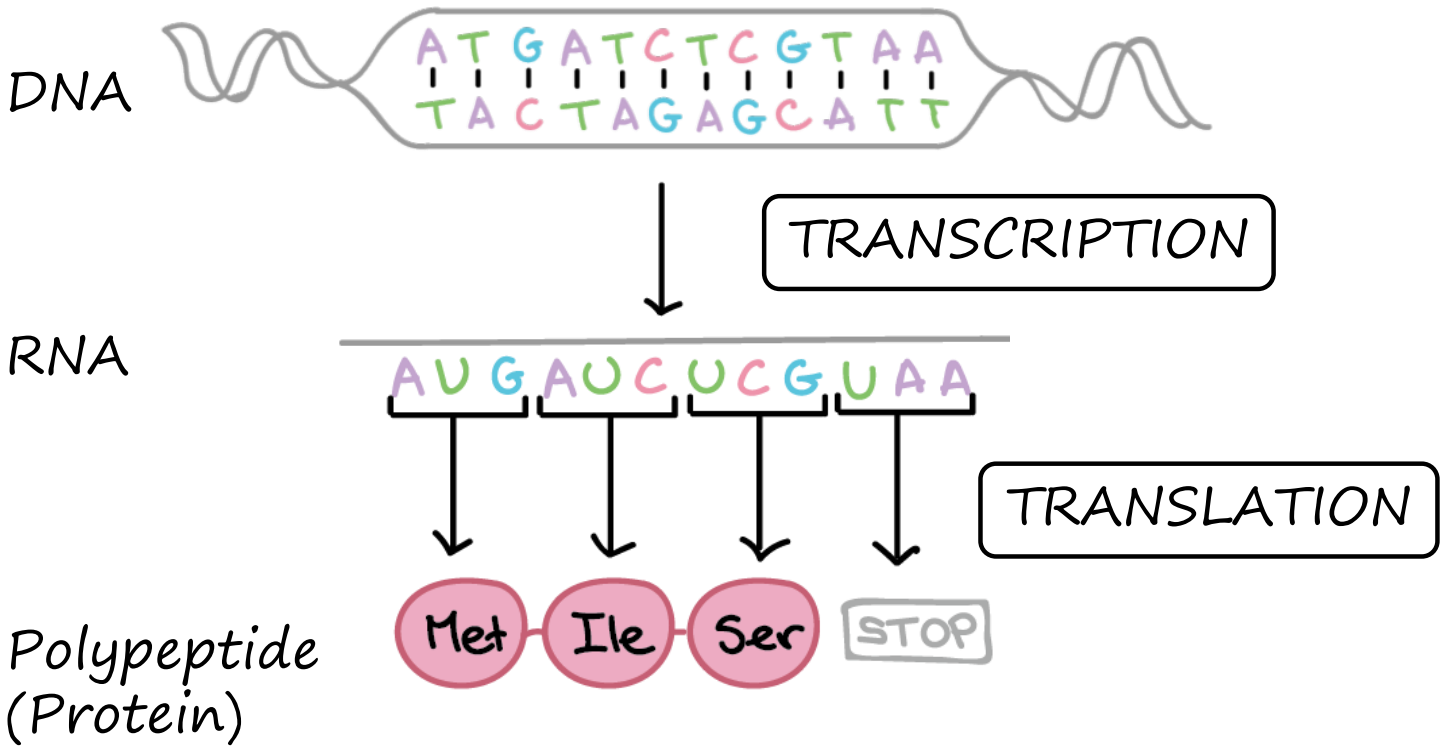
High school biology refresher

DNA

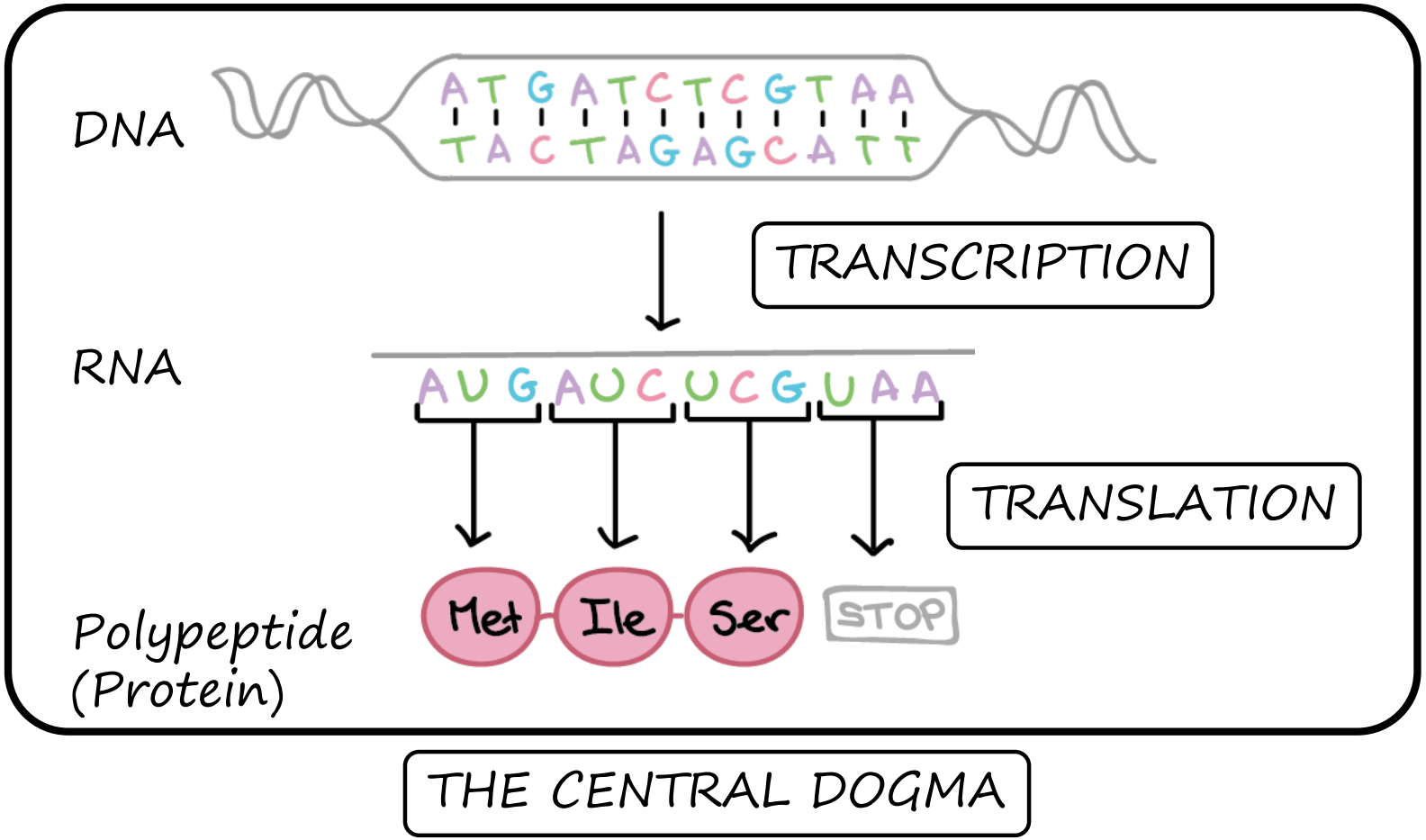
High school biology refresher



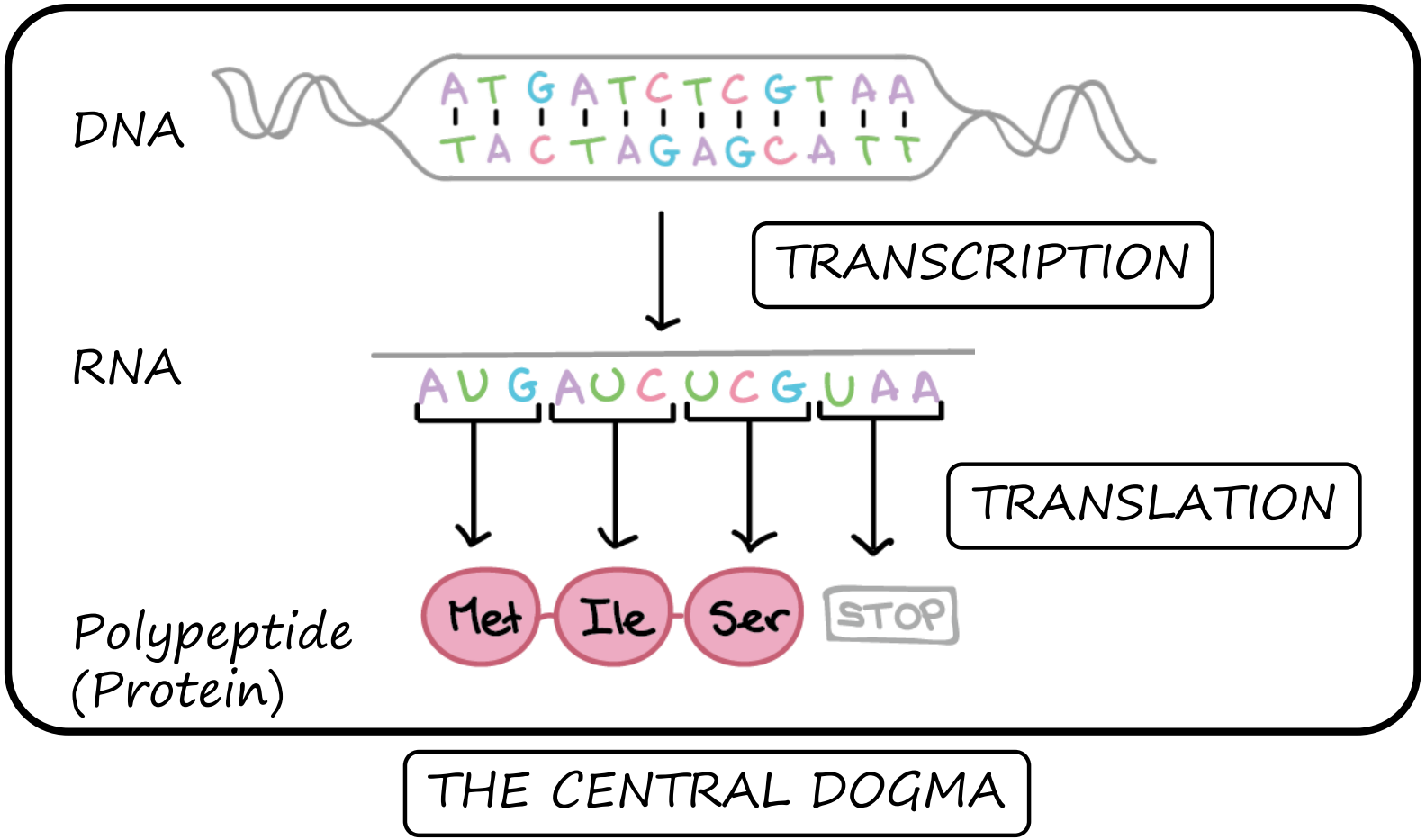
High school biology refresher



High school biology refresher



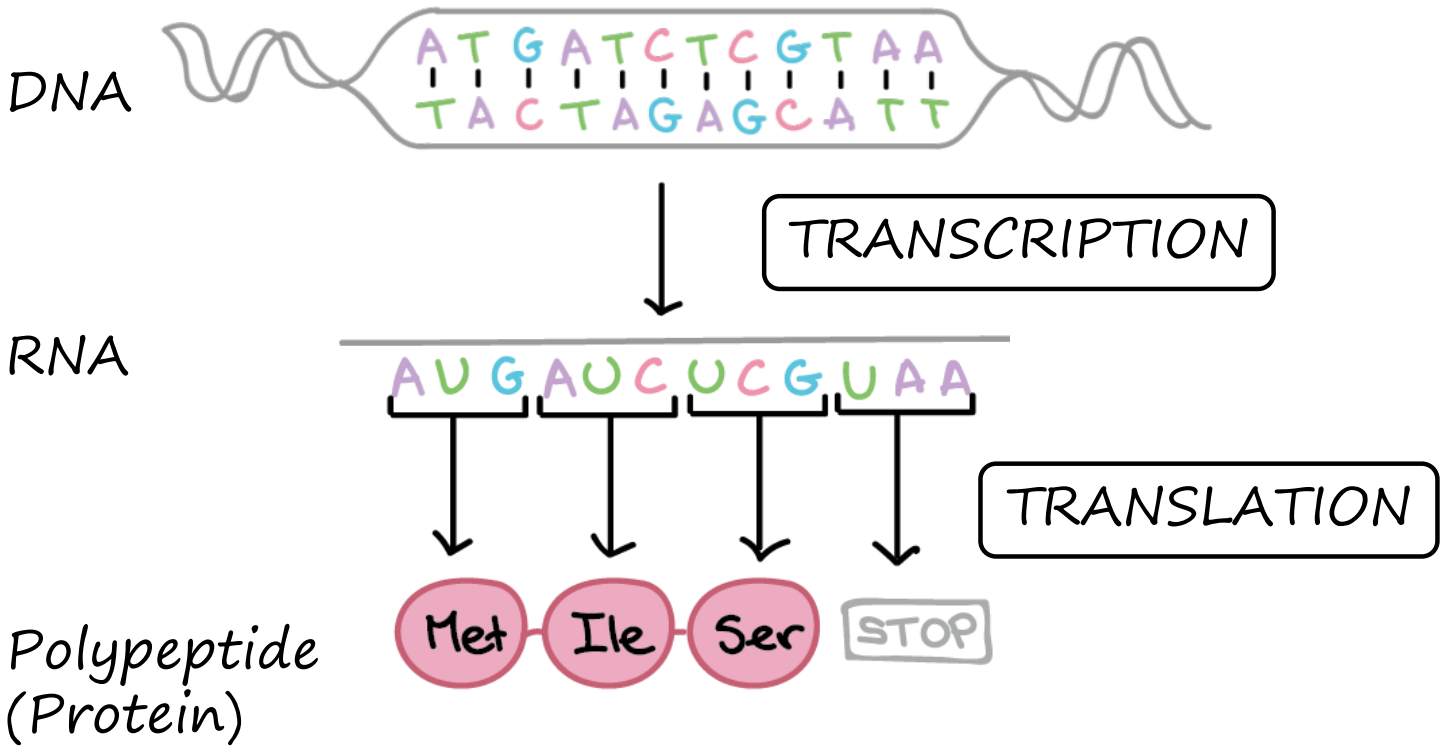
High school biology refresher



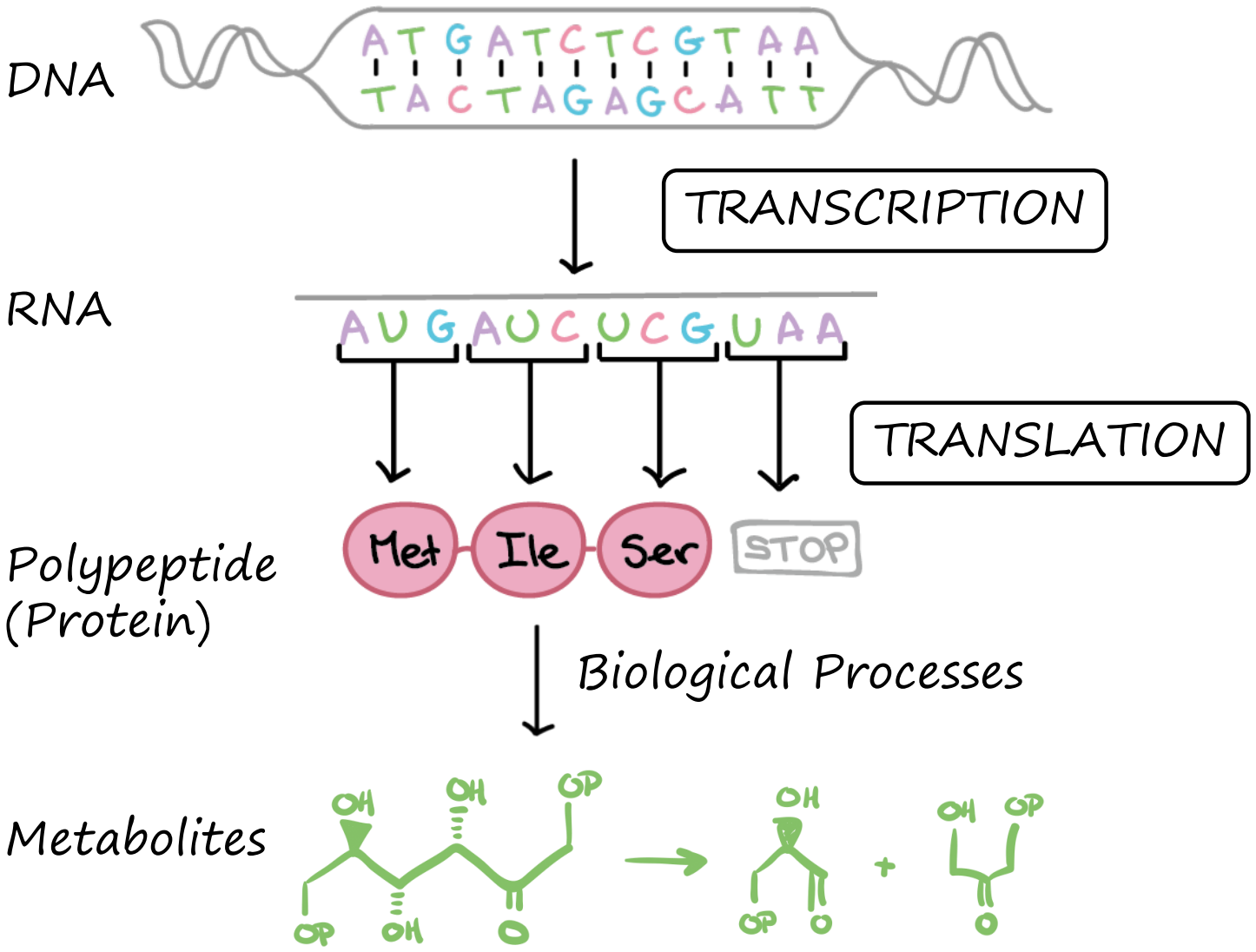
THE CENTRAL DOGMA

Classical definition of "gene": A DNA sequence that codes for RNA which is translated to a polypeptide chain.

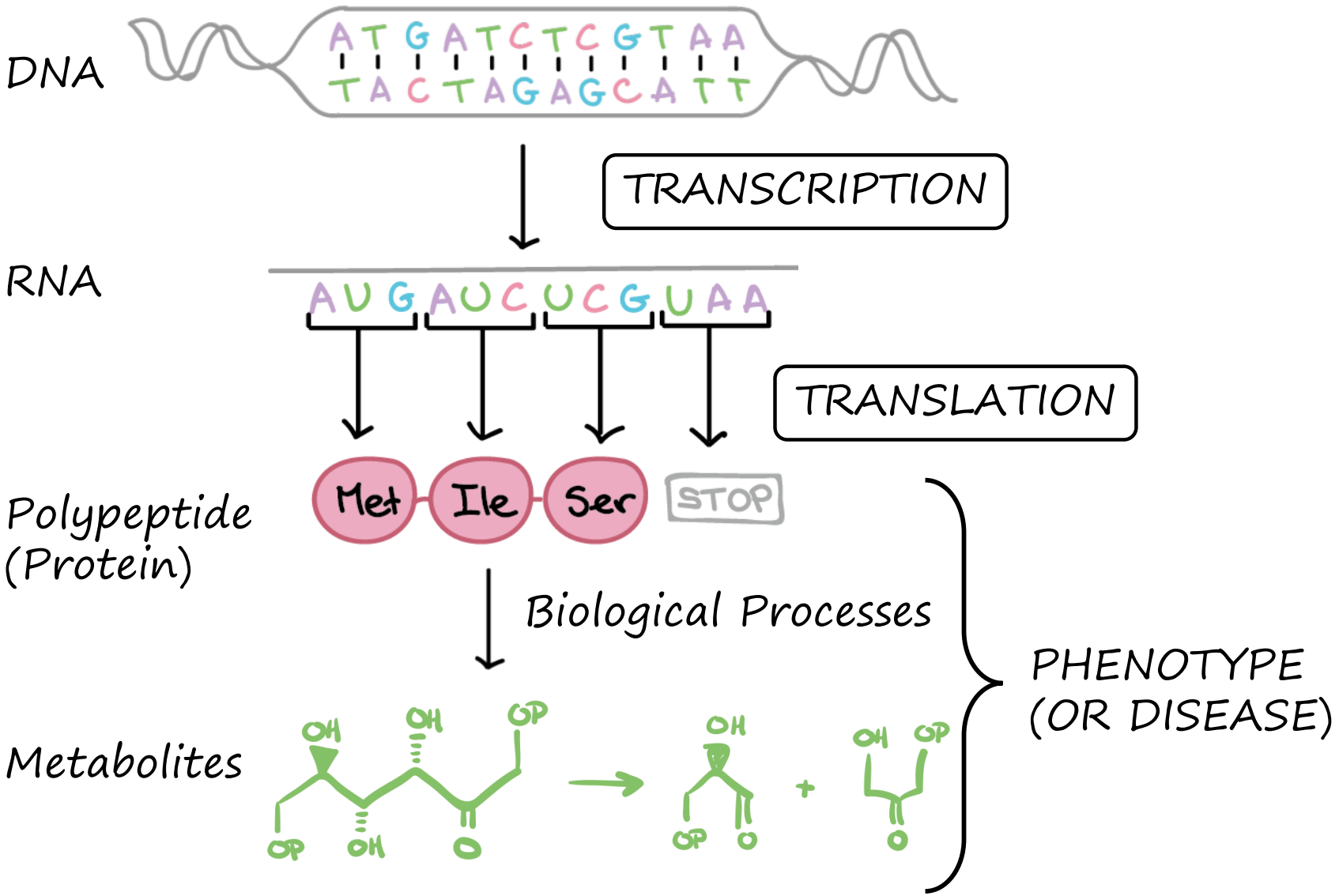
High school biology refresher



High school biology refresher



High school biology refresher



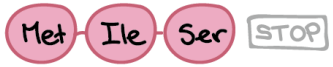
Overall Data Types

What it's called		Distribution in an individual
DNA 	Genomic → Germline	Same distribution across all tissues/cells
	Genomic → Somatic	Specific to the tissue/cell line, disease, age

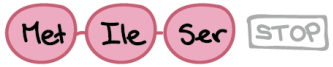
Overall Data Types

	What it's called		Distribution in an individual
DNA 	Genomic	Germline	Same distribution across all tissues/cells
		Somatic	Specific to the tissue/cell line, disease, age
RNA 	Transcriptomic		Specific to the tissue/cell line, exposure, disease and time point

Overall Data Types

	What it's called	Distribution in an individual
DNA 	Genomic → Germline	Same distribution across all tissues/cells
	Genomic → Somatic	Specific to the tissue/cell line, disease, age
RNA 	Transcriptomic	Specific to the tissue/cell line, exposure, disease and time point
Polypeptide (Protein) 	Proteomic	Specific to the tissue/cell line, exposure, disease and time point

Overall Data Types

	What it's called	Distribution in an individual
DNA 	Genomic → Germline	Same distribution across all tissues/cells
	Genomic → Somatic	Specific to the tissue/cell line, disease, age
RNA 	Transcriptomic	Specific to the tissue/cell line, exposure, disease and time point
Polypeptide (Protein) 	Proteomic	Specific to the tissue/cell line, exposure, disease and time point
Metabolites 	Metabolomic	Specific to the tissue/cell line, exposure and time point

Human Genome Project

Human Genome Project

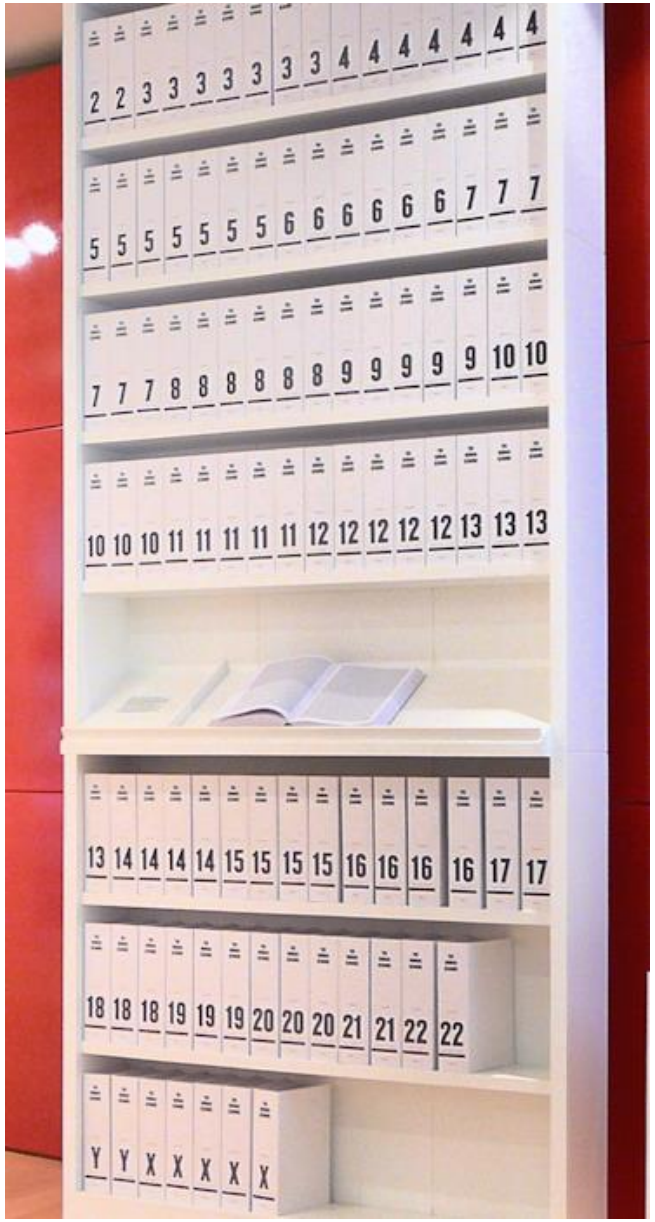
- 13-year project coordinated by the U.S. Department of Energy and the National Institutes 0f Health between 1999 and 2004.

Human Genome Project

- 13-year project coordinated by the U.S. Department of Energy and the National Institutes 0f Health between 1999 and 2004.
- The aim of the project was to map and identify all genes of the human genome -- to determine the coordinates of the human genome.

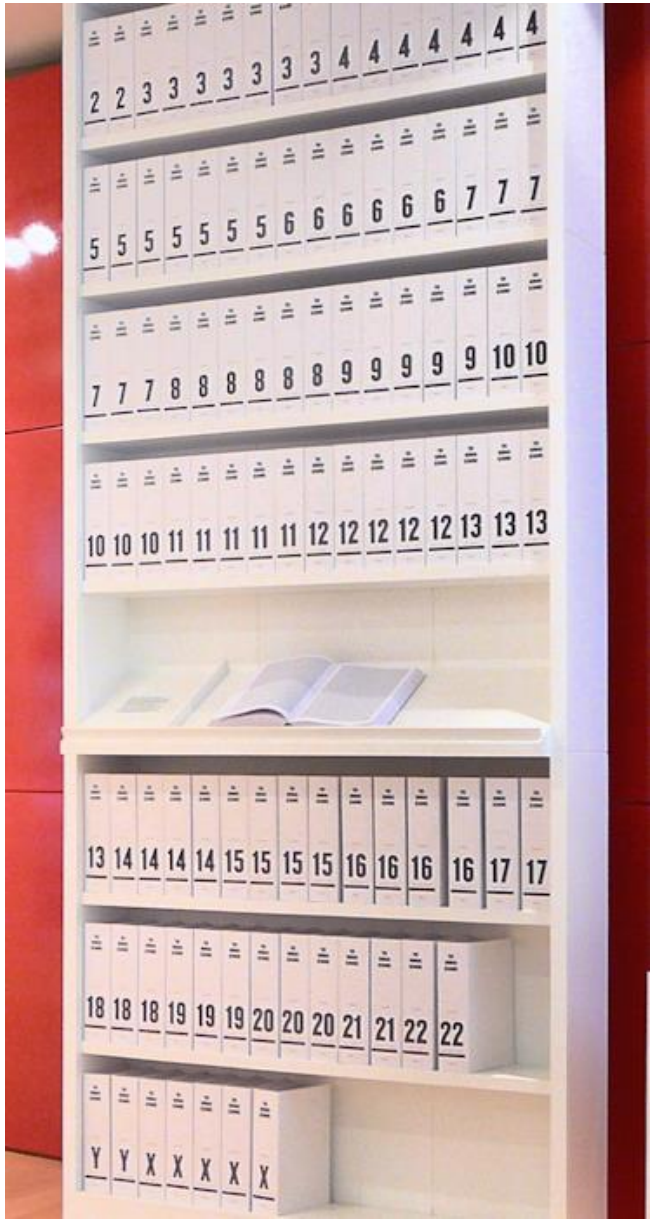
Human Genome Project

- 13-year project coordinated by the U.S. Department of Energy and the National Institutes 0f Health between 1999 and 2004.
- The aim of the project was to map and identify all genes of the human genome -- to determine the coordinates of the human genome.
- Project produced the first human reference genome, entirely changing genetics research.



Human Genome Project

- 13-year project coordinated by the U.S. Department of Energy and the National Institutes 0f Health between 1999 and 2004.
- The aim of the project was to map and identify all genes of the human genome -- to determine the coordinates of the human genome.
- Project produced the first human reference genome, entirely changing genetics research.
- New builds of the human reference genome gets released every several years further filling the gaps and correcting sequence mistakes.



What is genetic variation?

- Differences in DNA content or structure among individuals
  - Any two individuals have ~99.5% identical DNA.
- But the human genome is big:
  There are >100,000,000 known genetic variants in the human genome

~99.5% identical DNA
(differ at 1/ 620 - 1/750 bp)

~ 99% identical DNA

Aaron Quinlan Applied Computational Genomics, Lecture 05, https://github.com/quinlan-lab/applied-computational-genomics

Types of genetic variation

CTCCGAG
CTCTGAG

Single-nucleotide polymorphisms (**SNPs**)

"DNA spelling differences"

CTC--AG
CTCTGAG

Insertion/deletion polymorphisms (**INDELs**)

"Extra or missing DNA"

Structural variants (**SVs**)

"Large blocks of extra, missing or rearranged DNA"

Aaron Quinlan Applied Computational Genomics, Lecture 05, https://github.com/quinlan-lab/applied-computational-genomics

Types of genetic variation

CTCCGAG
CTCTGAG

Single-nucleotide polymorphisms (**SNPs**)

"DNA spelling differences"

CTC--AG
CTCTGAG

Insertion/deletion polymorphisms (**INDELs**)

"Extra or missing DNA"

Structural variants (**SVs**)

"Large blocks of extra, missing or rearranged DNA"

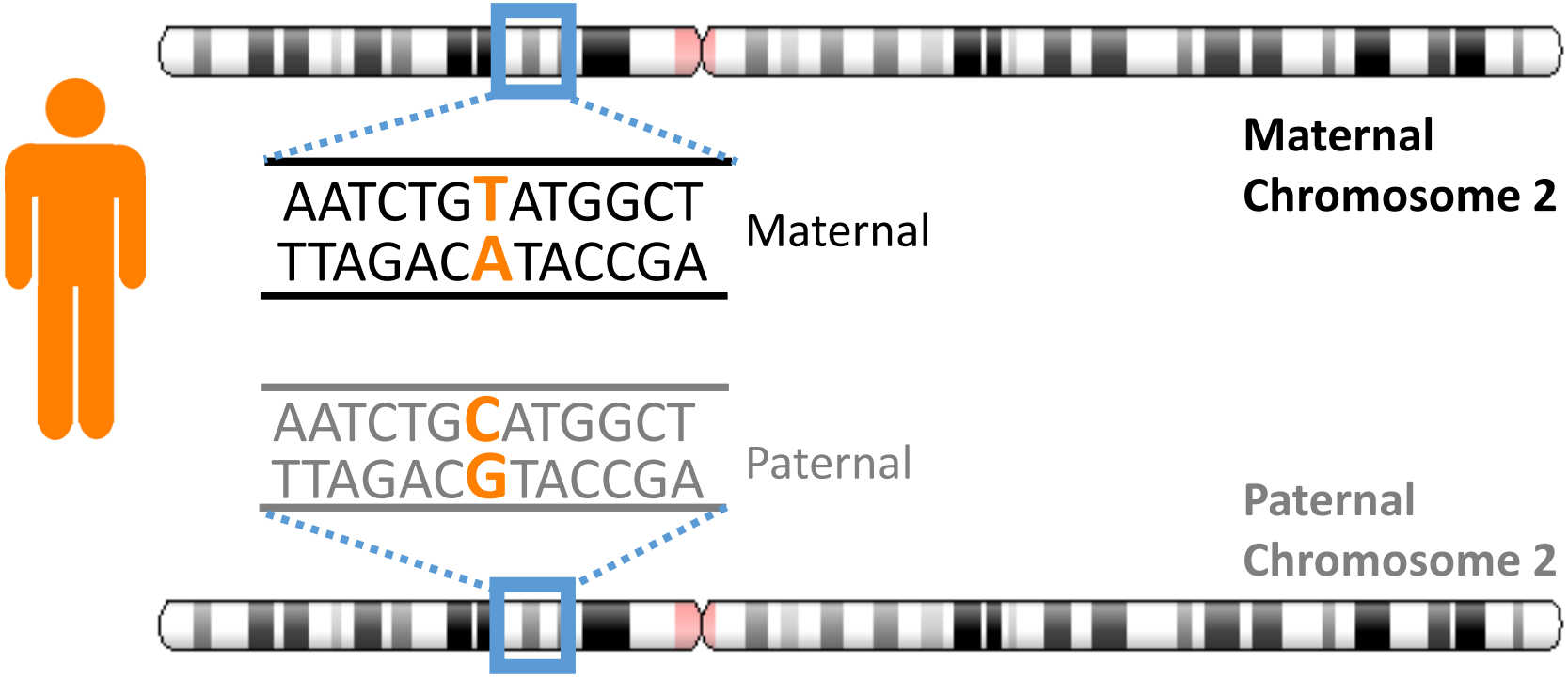
23andMe

ancestry™

Aaron Quinlan Applied Computational Genomics, Lecture 05, https://github.com/quinlan-lab/applied-computational-genomics

Single nucleotide polymorphism - SNP

A variation in a single DNA base pair that occurs at a specific position in the genome, where the **minor allele frequency of the variation is 1% or higher**.



Single nucleotide polymorphism - SNP

A variation in a single DNA base pair that occurs at a specific position in the genome, where the **minor allele frequency of the variation is 1% or higher**.

AATCTG**T**ATGGCT
TTAGAC**A**TACCGA

Maternal

Maternal Chromosome 2

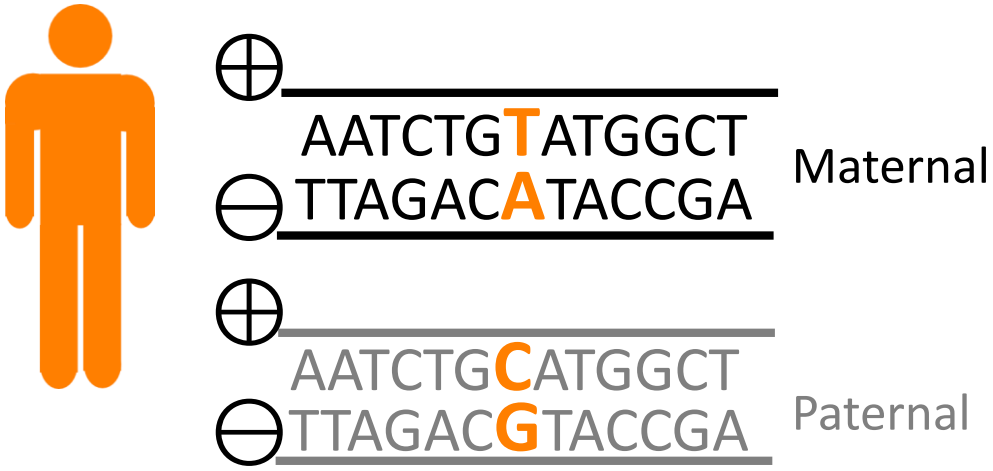
AATCTG**C**ATGGCT
TTAGAC**G**TACCGA

Paternal

Paternal Chromosome 2

Single nucleotide polymorphism - SNP

A variation in a single DNA base pair that occurs at a specific position in the genome, where the **minor allele frequency of the variation is 1% or higher**.

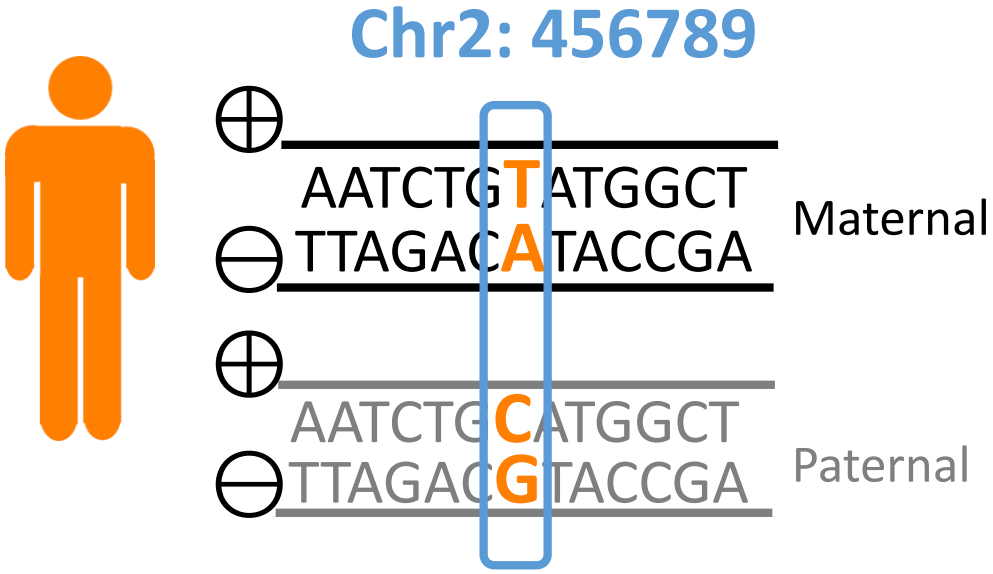


Maternal Chromosome 2

Paternal Chromosome 2

Single nucleotide polymorphism - SNP

A variation in a single DNA base pair that occurs at a specific position in the genome, where the **minor allele frequency of the variation is 1% or higher**.

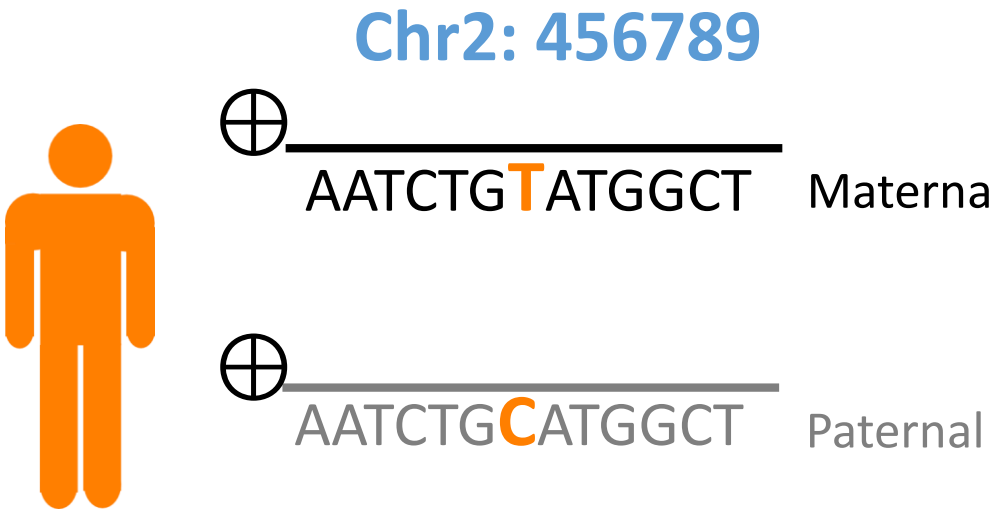


Maternal Chromosome 2

Paternal Chromosome 2

Single nucleotide polymorphism - SNP

A variation in a single DNA base pair that occurs at a specific position in the genome, where the **minor allele frequency of the variation is 1% or higher**.

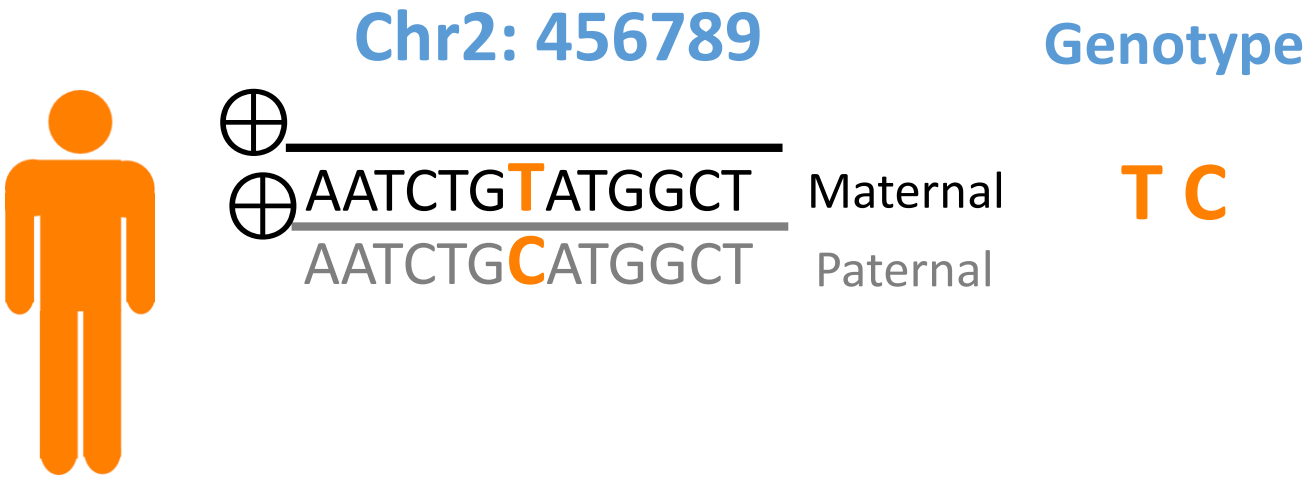


Single nucleotide polymorphism - SNP

A variation in a single DNA base pair that occurs at a specific position in the genome, where the **minor allele frequency of the variation is 1% or higher**.

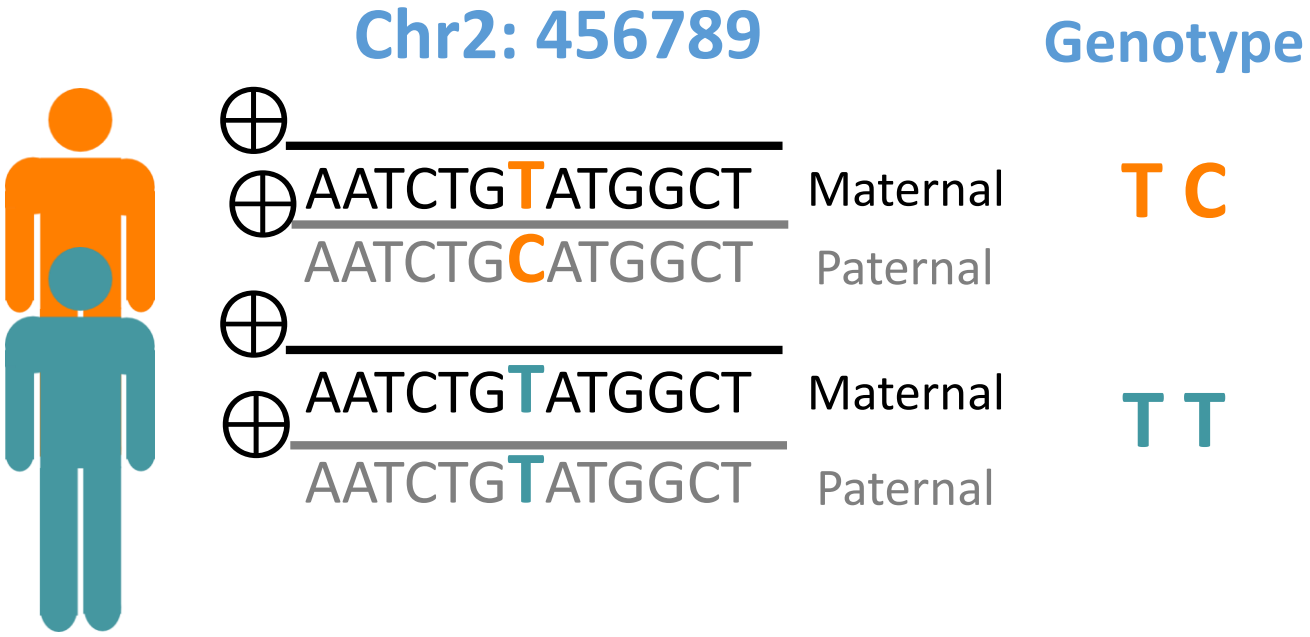
Single nucleotide polymorphism - SNP

A variation in a single DNA base pair that occurs at a specific position in the genome, where the **minor allele frequency of the variation is 1% or higher**.



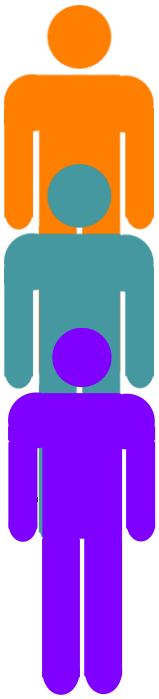
Single nucleotide polymorphism - SNP

A variation in a single DNA base pair that occurs at a specific position in the genome, where the **minor allele frequency of the variation is 1% or higher**.



Single nucleotide polymorphism - SNP

A variation in a single DNA base pair that occurs at a specific position in the genome, where the **minor allele frequency of the variation is 1% or higher**.



Chr2: 456789 **Genotype**

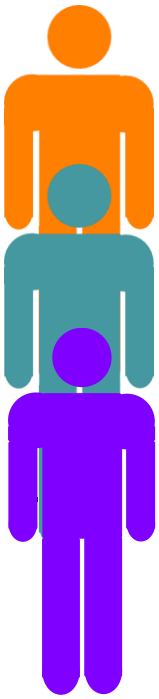
AATCTGTATGGCT Maternal
AATCTGCATGGCT Paternal **T C**

AATCTGTATGGCT Maternal
AATCTGTATGGCT Paternal **T T**

AATCTGCATGGCT Maternal
AATCTGCATGGCT Paternal **C C**

Single nucleotide polymorphism - SNP

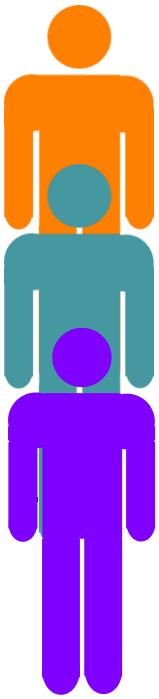
A variation in a single DNA base pair that occurs at a specific position in the genome, where the **minor allele frequency of the variation is 1% or higher**.



Chr2: 456789		Genotype	Allele Dosage of T
AATCTGTATGGCT	Maternal	T C	
AATCTGCATGGCT	Paternal		
AATCTGTATGGCT	Maternal	T T	
AATCTGTATGGCT	Paternal		
AATCTGCATGGCT	Maternal	C C	
AATCTGCATGGCT	Paternal		

Single nucleotide polymorphism - SNP

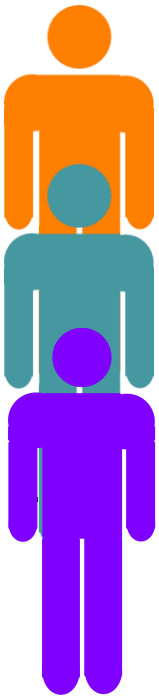
A variation in a single DNA base pair that occurs at a specific position in the genome, where the **minor allele frequency of the variation is 1% or higher**.



Chr2: 456789		Genotype	Allele Dosage of T
AATCTGTATGGCT	Maternal	T C	1
AATCTGCATGGCT	Paternal		
AATCTGTATGGCT	Maternal	T T	
AATCTGTATGGCT	Paternal		
AATCTGCATGGCT	Maternal	C C	
AATCTGCATGGCT	Paternal		

Single nucleotide polymorphism - SNP

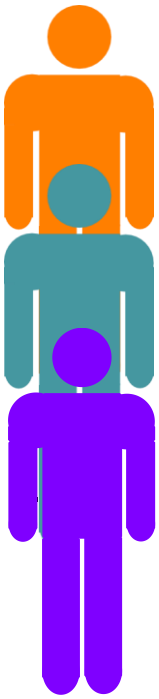
A variation in a single DNA base pair that occurs at a specific position in the genome, where the **minor allele frequency of the variation is 1% or higher**.



Chr2: 456789		Genotype	Allele Dosage of T
AATCTGTATGGCT	Maternal	T C	1
AATCTGCATGGCT	Paternal		
AATCTGTATGGCT	Maternal	T T	2
AATCTGTATGGCT	Paternal		
AATCTGCATGGCT	Maternal	C C	
AATCTGCATGGCT	Paternal		

Single nucleotide polymorphism - SNP

A variation in a single DNA base pair that occurs at a specific position in the genome, where the **minor allele frequency of the variation is 1% or higher**.



Chr2: 456789		Genotype	Allele Dosage of T
AATCTGTATGGCT	Maternal	T C	1
AATCTGCATGGCT	Paternal		
AATCTGTATGGCT	Maternal	T T	2
AATCTGTATGGCT	Paternal		
AATCTGCATGGCT	Maternal	C C	0
AATCTGCATGGCT	Paternal		

Haplotype

- A cluster of single nucleotide polymorphisms (SNPs) on the same chromosome that often inherited together (due to linkage disequilibrium).
- Linkage disequilibrium (LD) is the non-random inheritance (occurrence in population) of alleles at different loci.

	Haplotypes
GGTTCAATCTGTATGGCTAAA	T T A
GGGTCAATCTGCATGGCTACA	G C C
GGTTCAATCTGTATGGCTAAA	T T A
GGTTCAATCTGTATGGCTAAA	T T A

Haplotype

- A cluster of single nucleotide polymorphisms (SNPs) on the same chromosome that often inherited together (due to linkage disequilibrium).
- Linkage disequilibrium (LD) is the non-random inheritance (occurrence in population) of alleles at different loci.

Haplotypes

T T A
G C C

T T A
T T A

Haplotype

- A cluster of single nucleotide polymorphisms (SNPs) on the same chromosome that often inherited together (due to linkage disequilibrium).
- Linkage disequilibrium (LD) is the non-random inheritance (occurrence in population) of alleles at different loci.

Haplotypes

T T A
G C C

T T A
T T A

Reference Alleles

T T C

Haplotype

- A cluster of single nucleotide polymorphisms (SNPs) on the same chromosome that often inherited together (due to linkage disequilibrium).
- Linkage disequilibrium (LD) is the non-random inheritance (occurrence in population) of alleles at different loci.

Haplotypes	Reference Alleles	Allele dosages of Haplotypes
T T A		1 1 0
G C C	T T C	0 0 1
T T A		1 1 0
T T A		1 1 0

Global reference for human genetic variation

Over 88 million variants:
84.7 million SNPs
3.6 million Indels

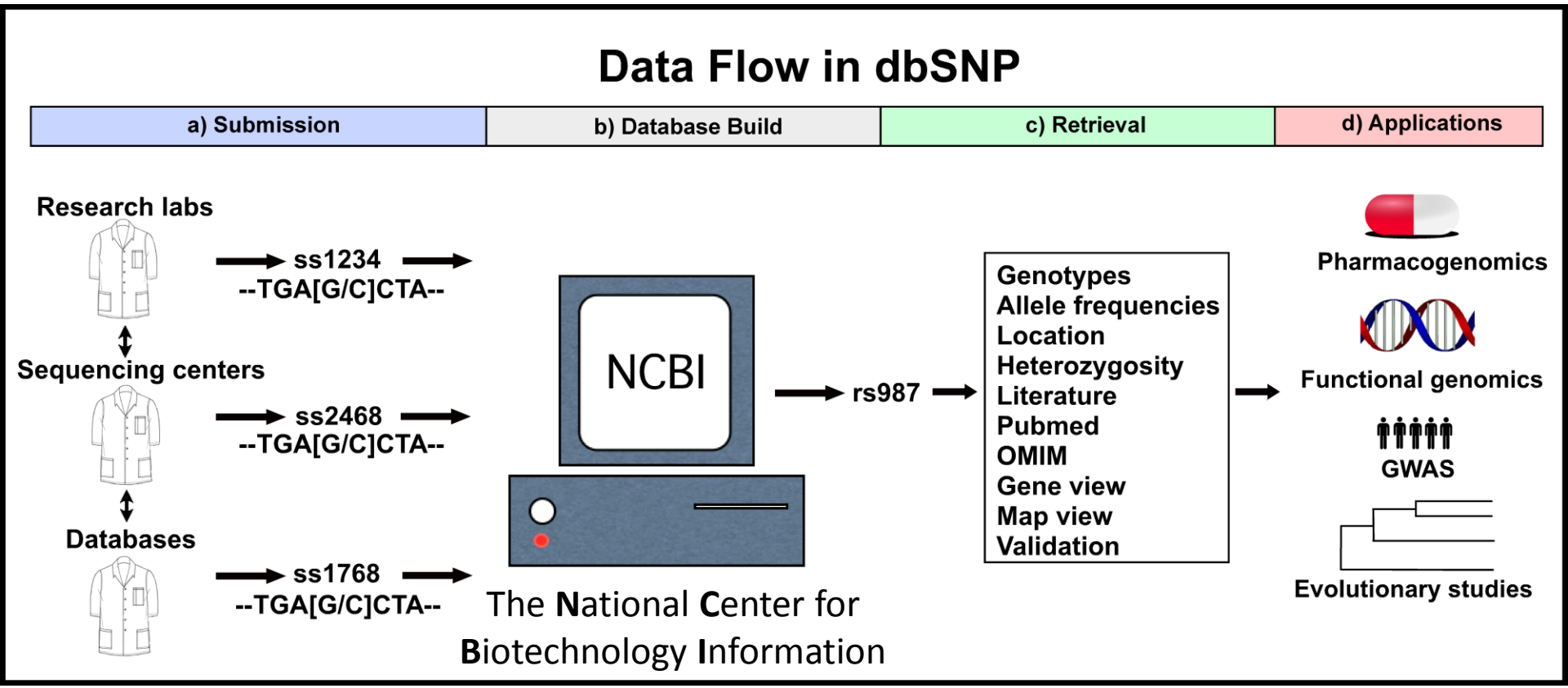
HapMap Project

1.6 million SNPs
+ CNVs

1000 Genome Project

The International Genome Sample Resource (IGSR) was established to ensure the ongoing usability of data generated by the 1000 Genomes Project and to extend the data set. More information is available about the IGSR.

Single Nucleotide Polymorphism Database (dbSNP)



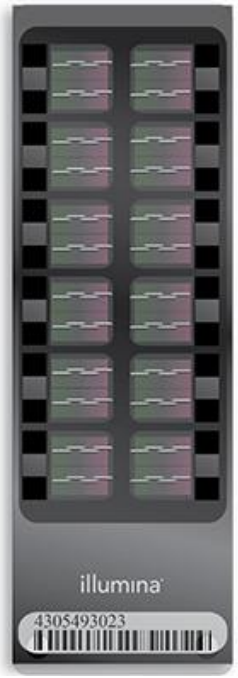
by Martin MacInnis and Greg Baute.

# Submissions (ss)	# RefSNP	# of rs in gene
538,341,120	150,482,731	87,339,846

From Nov 24, 2015

Identifying human genetic variation on a large scale

SNP Arrays (SNP-Chips)

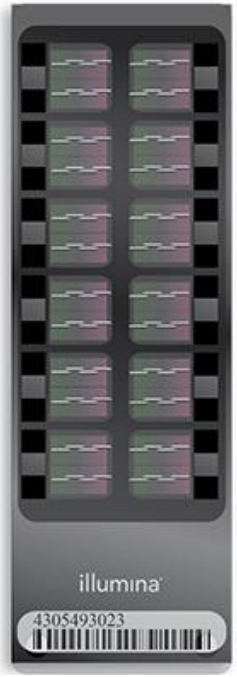


Illumina Human OmniExpress Bead Chip

Next Generation Sequencing

Illumina HiSeq 3000

Identifying human genetic variation on a large scale



Illumina Human OmniExpress Bead Chip

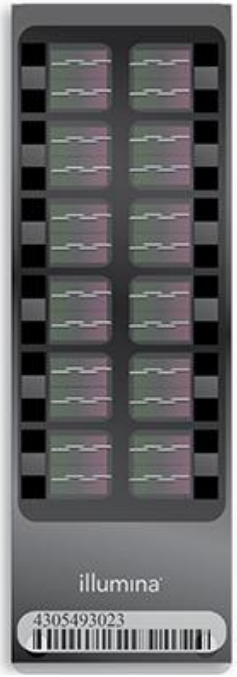
SNP Arrays (SNP-Chips)

- Can genotype ~750K SNPs.
- With genotype imputation it can yield up to ~30M SNPs.
- Large consortia data and LD structure is used for chip design.
- Cannot detect all types of genetic variation.
- Much cheaper than sequencing.
- Relatively easier data to work with.
- Frequently used for Genome Wide Association Studies (GWAS)

Next Generation Sequencing

Illumina HiSeq 3000

Identifying human genetic variation on a large scale



Illumina Human OmniExpress Bead Chip

SNP Arrays (SNP-Chips)

- Can genotype ~750K SNPs.
- With genotype imputation it can yield up to ~30M SNPs.
- Large consortia data and LD structure is used for chip design.
- Cannot detect all types of genetic variation.
- Much cheaper than sequencing.
- Relatively easier data to work with.
- Frequently used for Genome Wide Association Studies (GWAS)

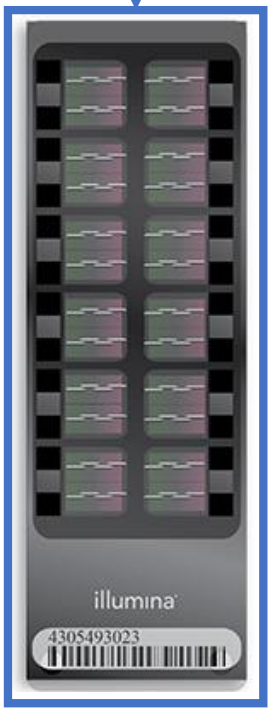
Next Generation Sequencing

- Does not depend on prior knowledge
- Can detect (almost) any type of genetic information.

Illumina HiSeq 3000

- Generated data is more difficult to work with (compared to SNP arrays).
- 3-4 times more expensive than SNP Arrays.
- Making it less desirable for large scale (thousands of samples) genomic studies.

Identifying human genetic variation on a large scale



Illumina Human OmniExpress Bead Chip

SNP Arrays (SNP-Chips)

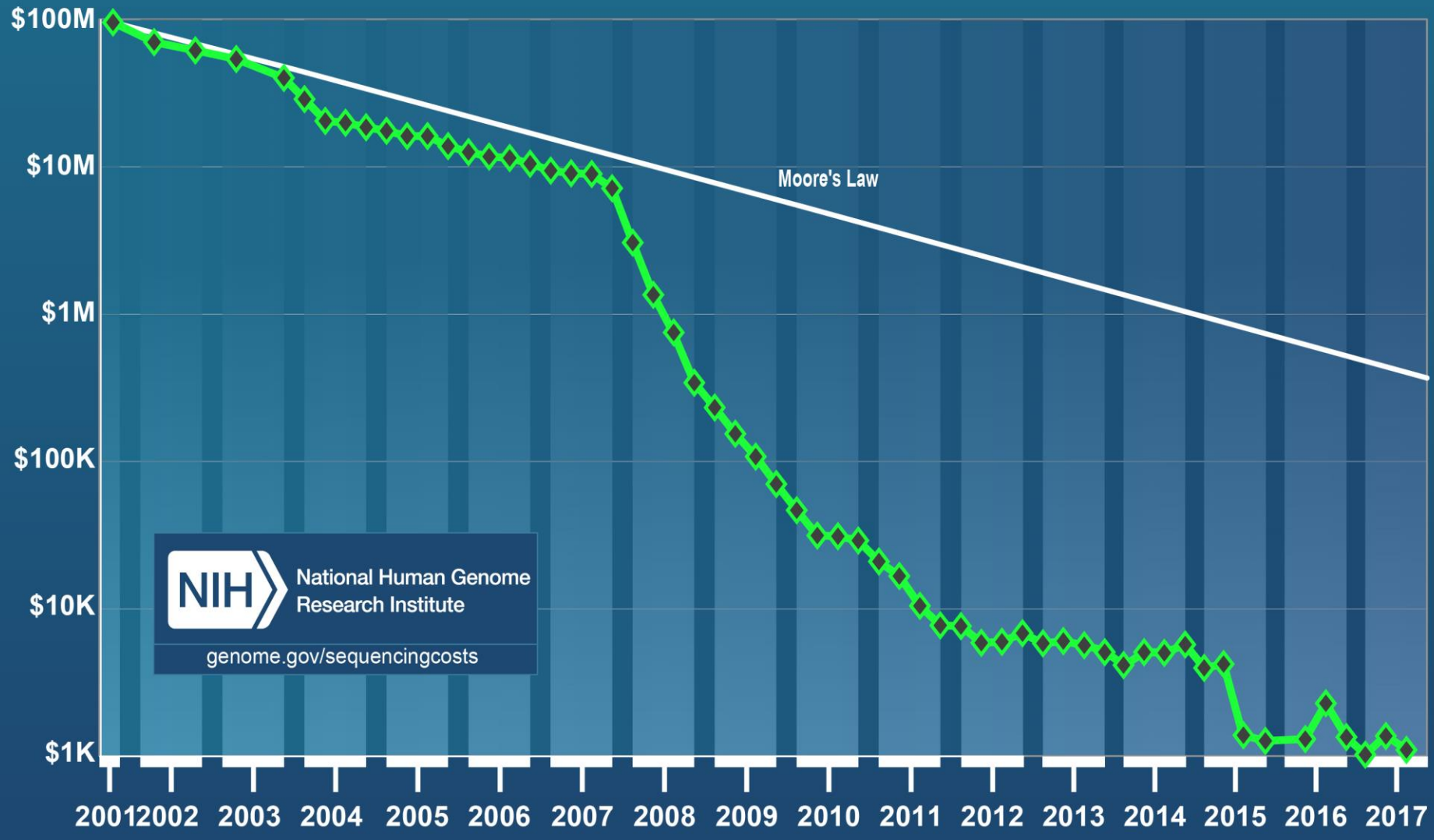
- Can genotype ~750K SNPs.
- With genotype imputation it can yield up to ~30M SNPs.
- Large consortia data and LD structure is used for chip design.
- Cannot detect all types of genetic variation.
- Much cheaper than sequencing.
- Relatively easier data to work with.
- Frequently used for Genome Wide Association Studies (GWAS)

Next Generation Sequencing

Illumina HiSeq 3000

- Does not depend on prior knowledge
- Can detect (almost) any type of genetic information.
- Generated data is more difficult to work with (compared to SNP arrays).
- 3-4 times more expensive than SNP Arrays.
- Making it less desirable for large scale (thousands of samples) genomic studies.

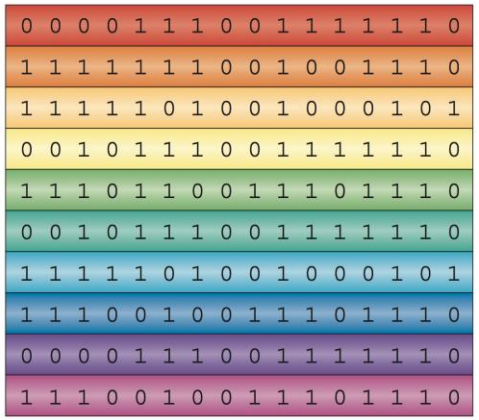
Cost per Genome



Genotype imputation: Statistical inference of unobserved genotypes.

Columns:
SNP positions

Rows:
Samples (Patients)

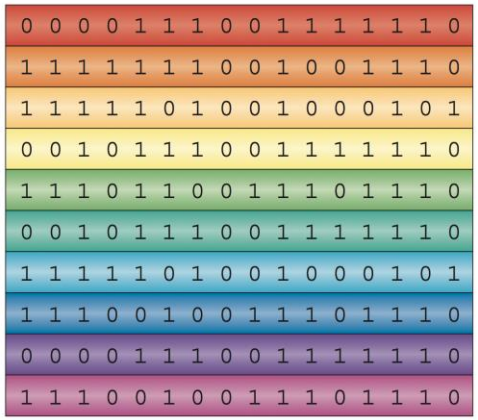


Marchini, Jonathan, and Bryan Howie.
Nature Reviews Genetics 11.7 (2010): 499-511.

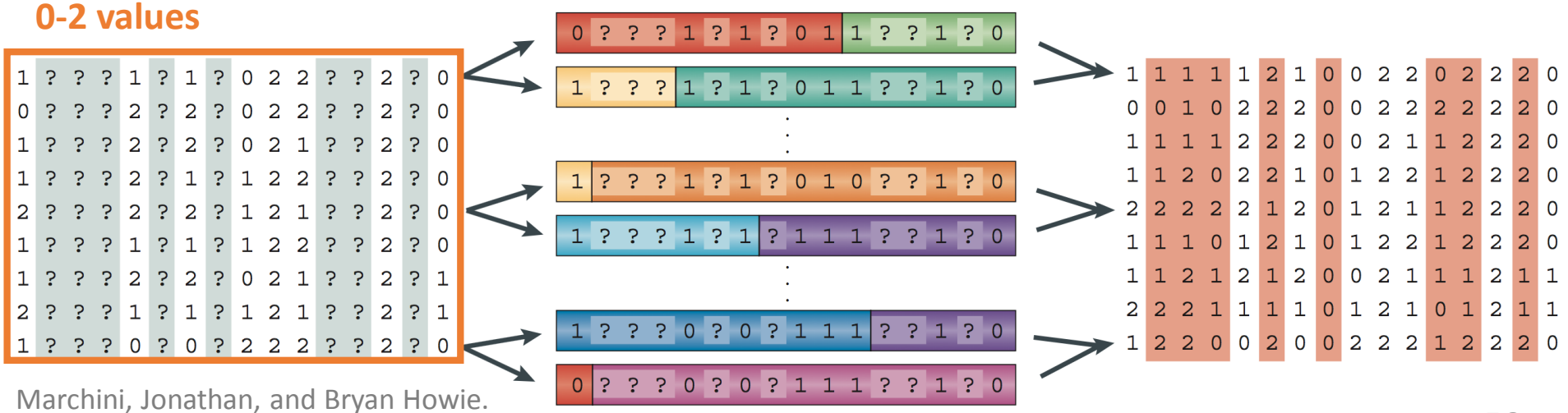
Genotype imputation: Statistical inference of unobserved genotypes.

Columns:
SNP positions

Rows:
Samples (Patients)

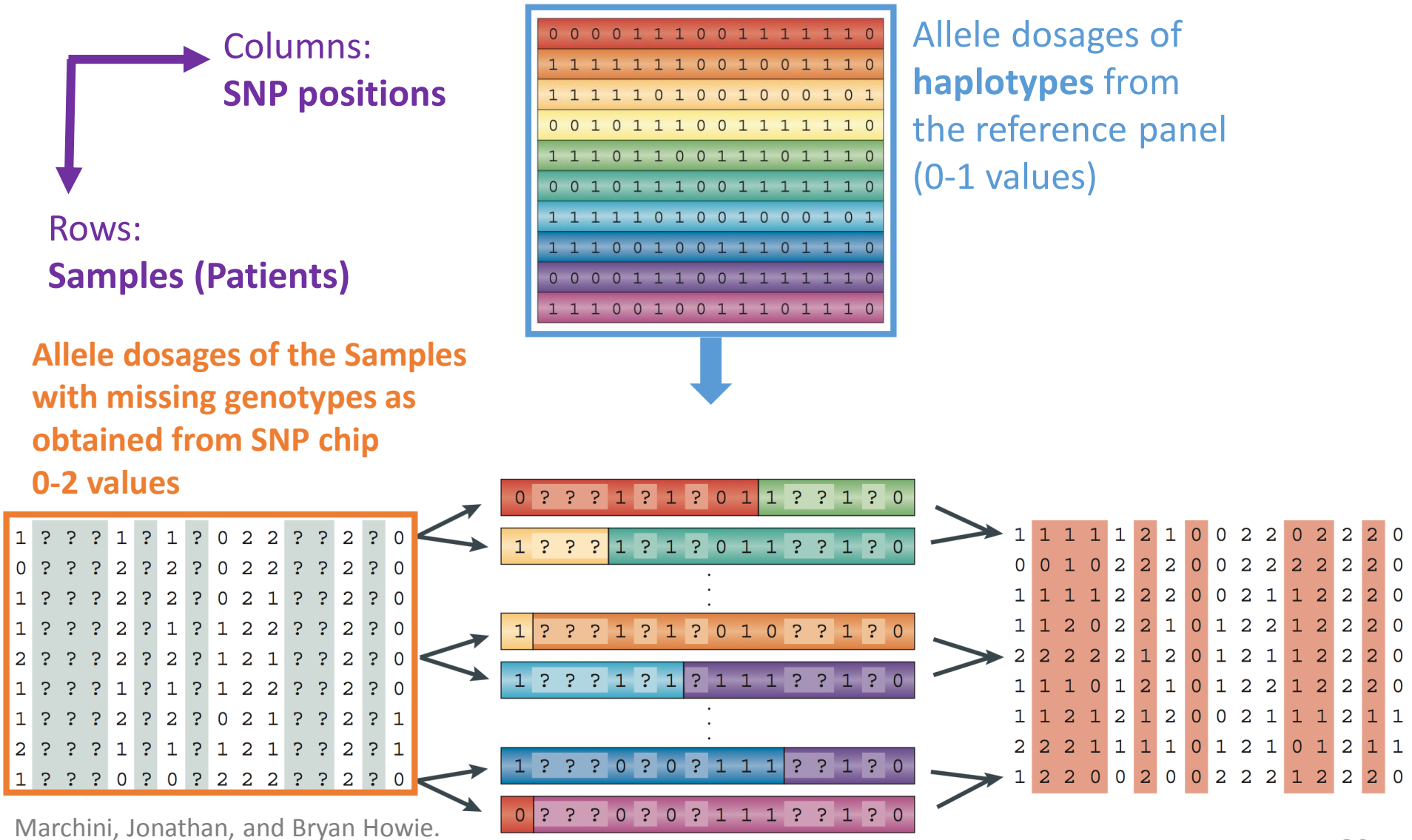


Allele dosages of the Samples with missing genotypes as obtained from SNP chip
0-2 values



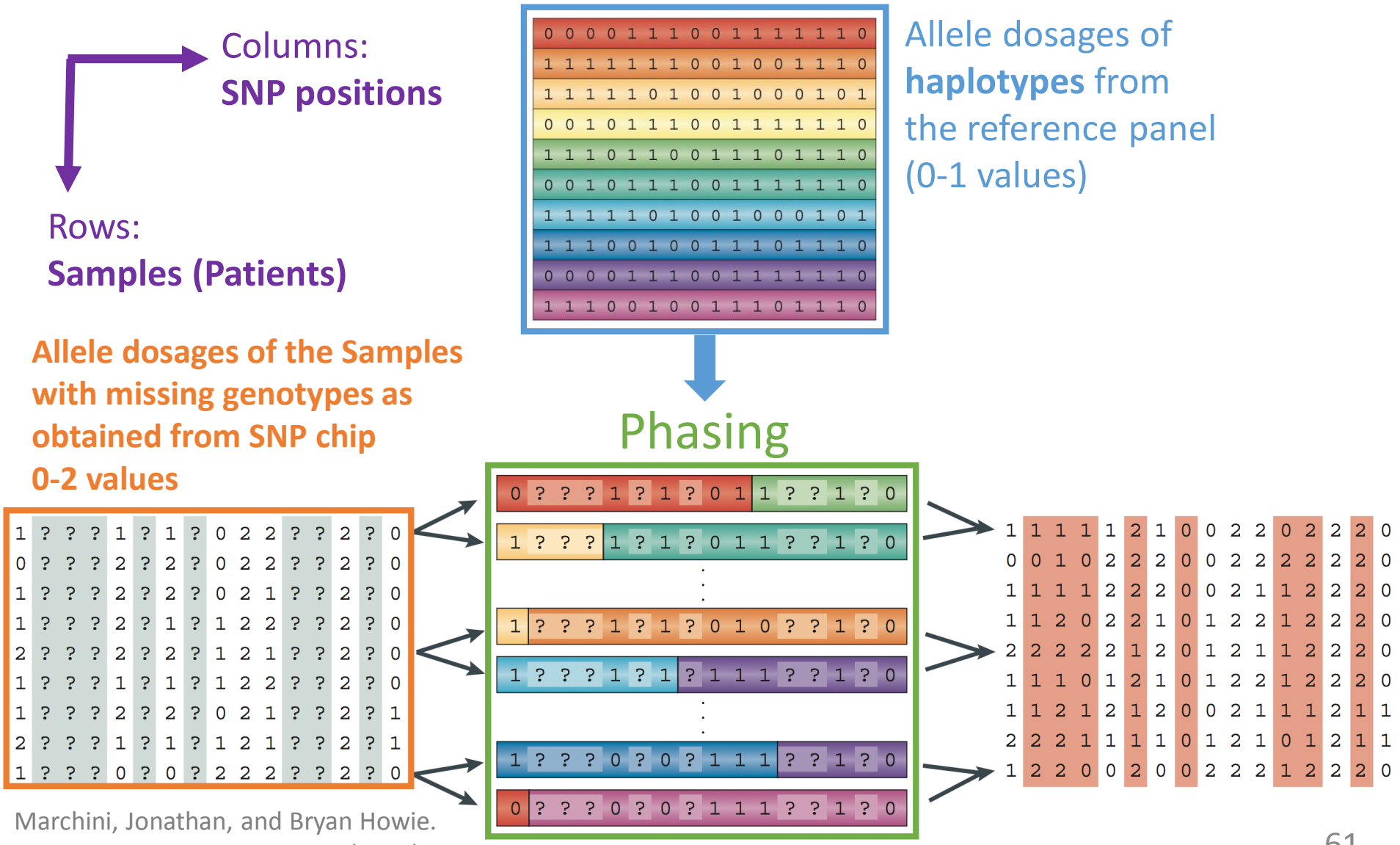
Marchini, Jonathan, and Bryan Howie.
Nature Reviews Genetics 11.7 (2010): 499-511.

Genotype imputation: Statistical inference of unobserved genotypes.



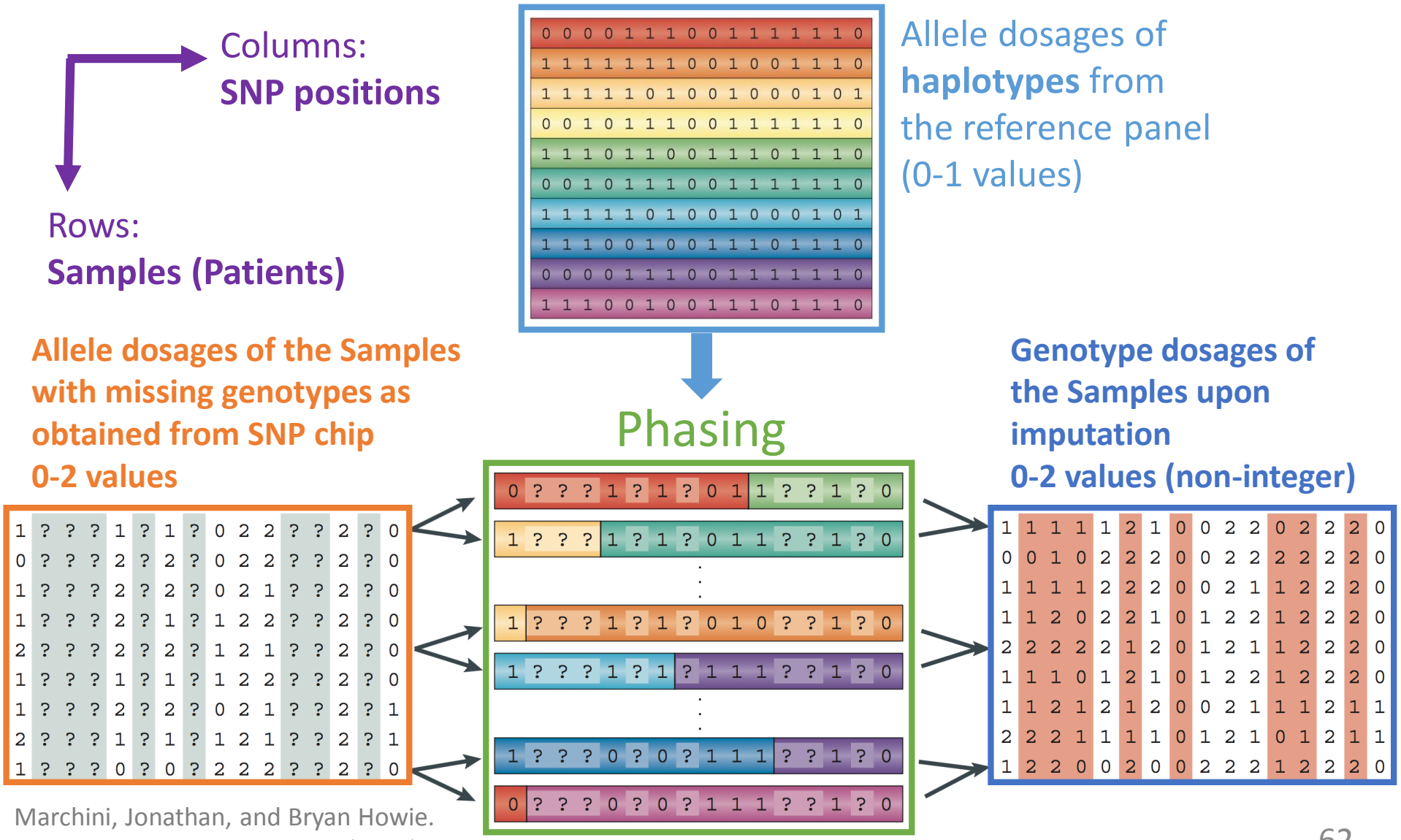
Marchini, Jonathan, and Bryan Howie. *Nature Reviews Genetics* 11.7 (2010): 499-511.

Genotype imputation: Statistical inference of unobserved genotypes.



Marchini, Jonathan, and Bryan Howie. *Nature Reviews Genetics* 11.7 (2010): 499-511.

Genotype imputation: Statistical inference of unobserved genotypes.



Marchini, Jonathan, and Bryan Howie. *Nature Reviews Genetics* 11.7 (2010): 499-511.

Imputation results

- Genotype dosage explained:
  For **a specific SNP** and a one individual

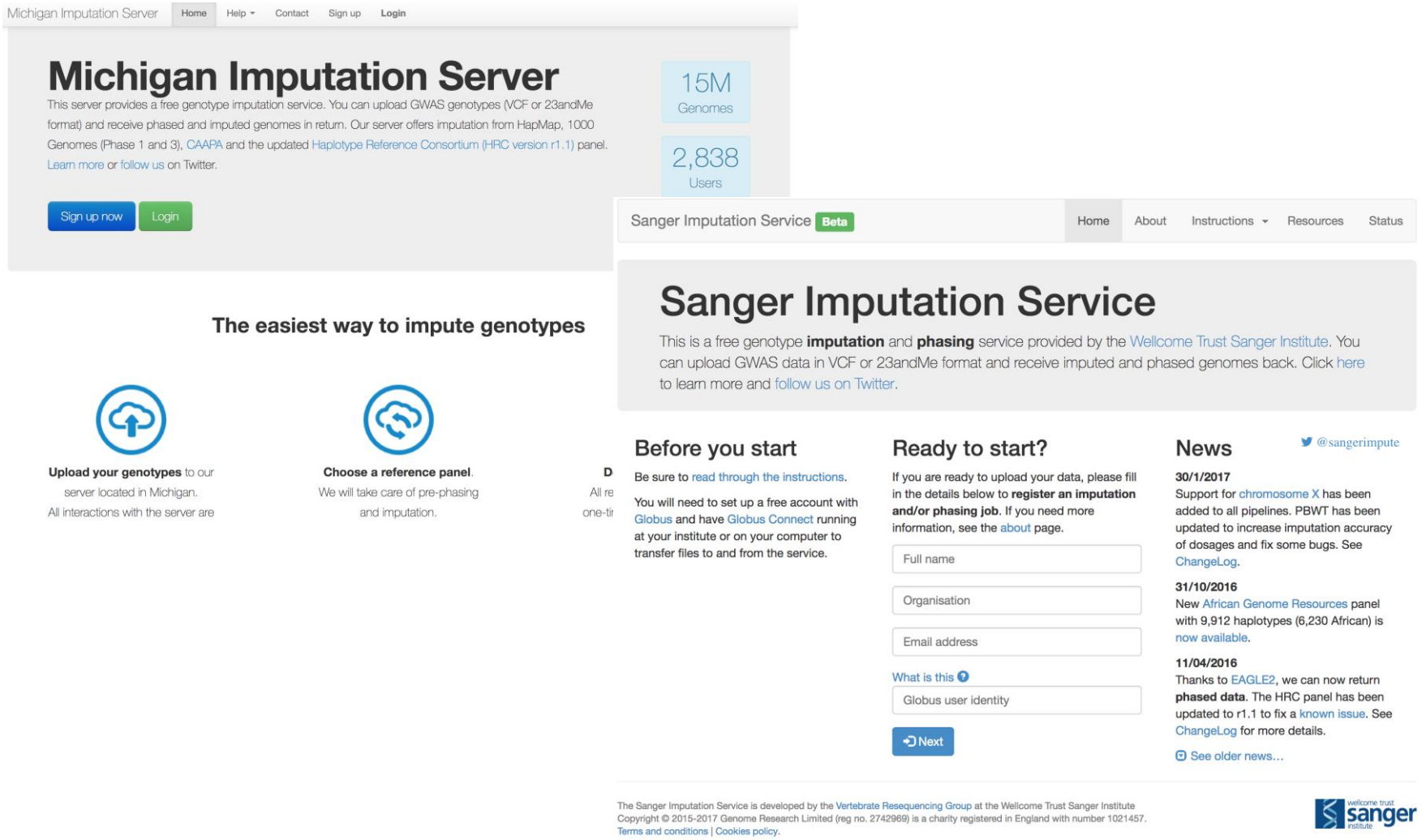
Estimated Allele Dosage

AA Aa aa

$$(0 \times 0.98) + (1 \times 0.01) + (2 \times 0.01) = 0.03$$

- Quality of estimation:
  Depending on the SNP density of the imputed region, estimation quality can differ. Regions with higher SNP density tend to give better estimations than low density regions.
- Info Score: (a score statistic computed during imputation) gives the measure measure of the observed statistical information associated with the allele frequency estimate
- Certainty: average certainty of best-guess genotypes

You can impute your genome too!



So how does **23andMe** or **ancestry** knows if you can smell a funny odor when you pee after eating asparagus?

Genome Wide Association Study (GWAS)

"An observational study of a genome-wide set of genetic variants in different individuals to see if any variant is associated with a trait." - Wikipedia

Genome Wide Association Study (GWAS)

"An observational study of a genome-wide set of genetic variants in different individuals to see if any variant is associated with a trait." - Wikipedia

Chi squared

$$\sum \frac{(f_o + f_e)^2}{f_o}$$

Genome Wide Association Study (GWAS)

"An observational study of a genome-wide set of genetic variants in different individuals to see if any variant is associated with a trait." - Wikipedia

Chi squared

$$\sum \frac{(f_o + f_e)^2}{f_o}$$

Simple linear regression

$$y_i = \beta_0 + \beta_1 x_i$$

Genome Wide Association Study (GWAS)

"An observational study of a genome-wide set of genetic variants in different individuals to see if any variant is associated with a trait." - Wikipedia

Chi squared

$$\sum \frac{(f_o + f_e)^2}{f_o}$$

Simple linear regression

$$y_i = \beta_0 + \beta_1 x_i$$

Multiple regression

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}$$

Genome Wide Association Study (GWAS)

"An observational study of a genome-wide set of genetic variants in different individuals to see if any variant is associated with a trait." - Wikipedia

Chi squared

$$\sum \frac{(f_o + f_e)^2}{f_o}$$

Simple linear regression

$$y_i = \beta_0 + \beta_1 x_i$$

Multiple regression

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}$$

Genome Wide Association Study (GWAS)

"An observational study of a genome-wide set of genetic variants in different individuals to see if any variant is associated with a trait." - Wikipedia

Chi squared

$$\sum \frac{(f_o + f_e)^2}{f_o}$$

Simple linear regression

$$y_i = \beta_0 + \beta_1 x_i$$

Multiple regression

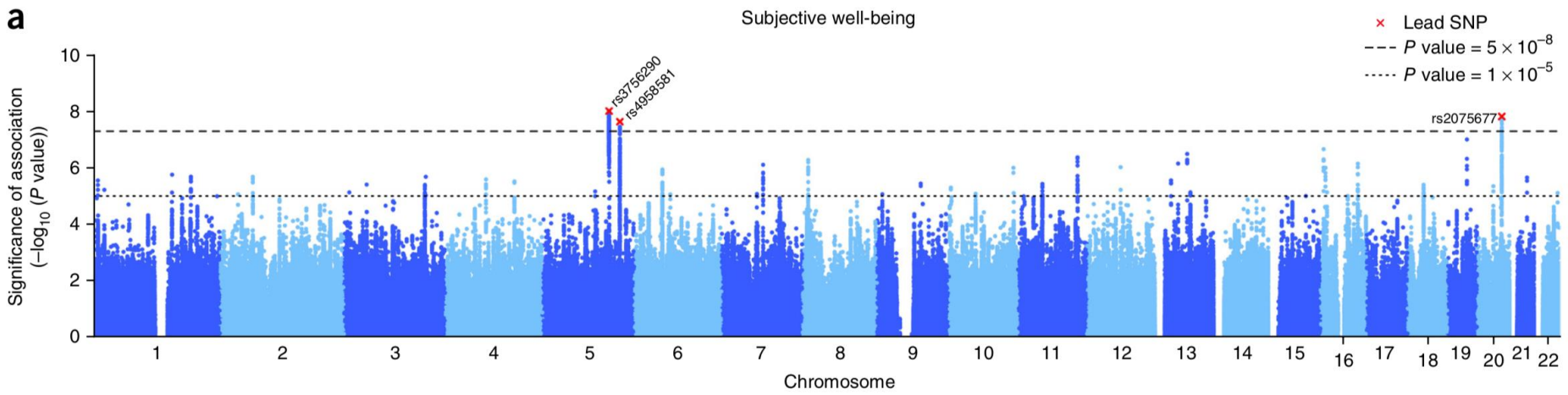
$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}$$

$$j = 1, \cdots, \sim 12 \times 10^6$$

nature genetics

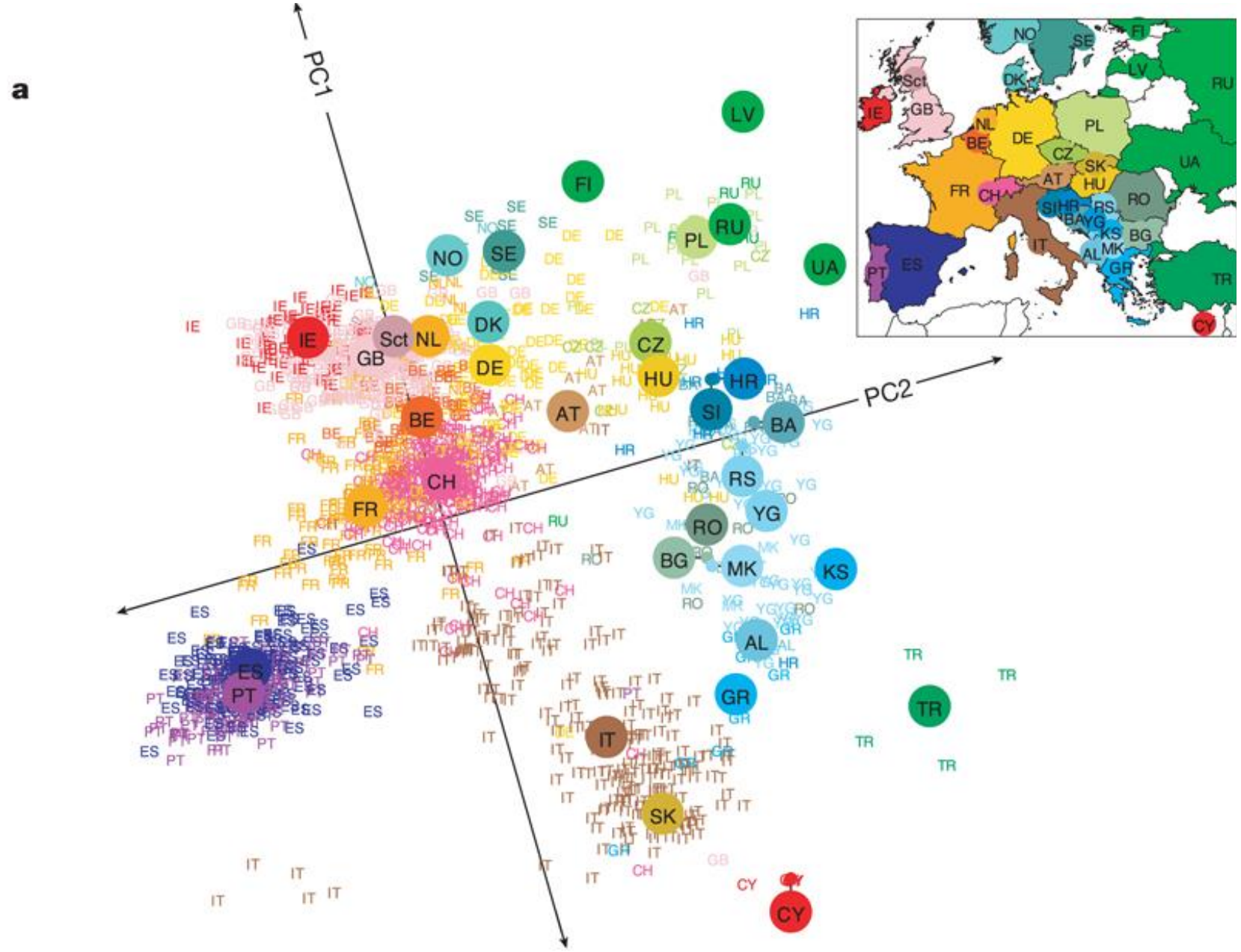
Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses

[…] conducted genome-wide association studies of three phenotypes: subjective well-being (***n* = 298,420**), depressive symptoms (***n* = 161,460**), and neuroticism (***n* = 170,911**).

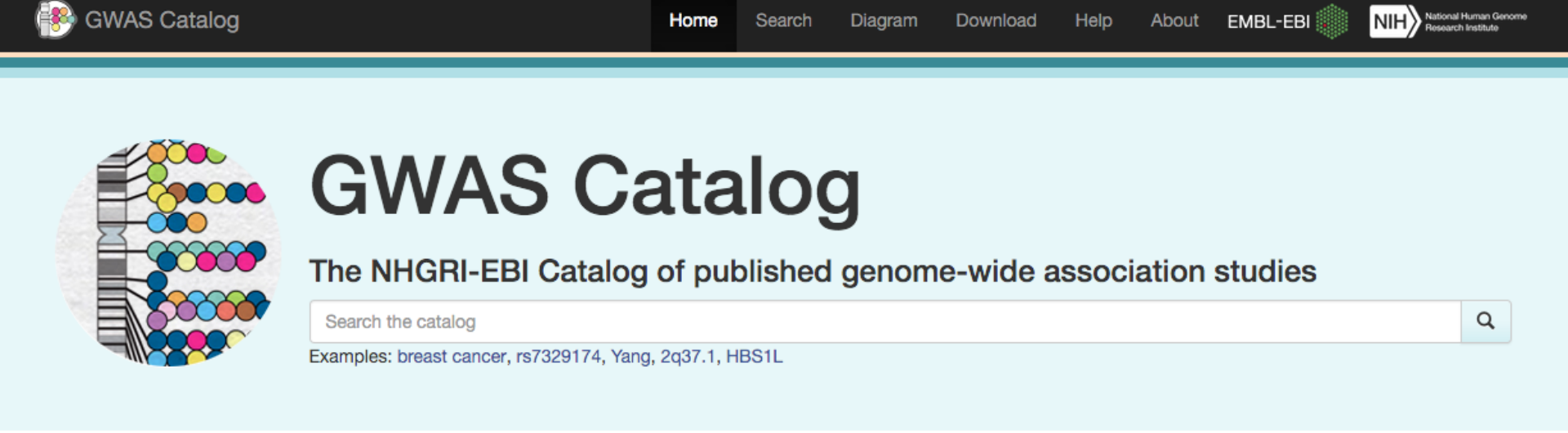


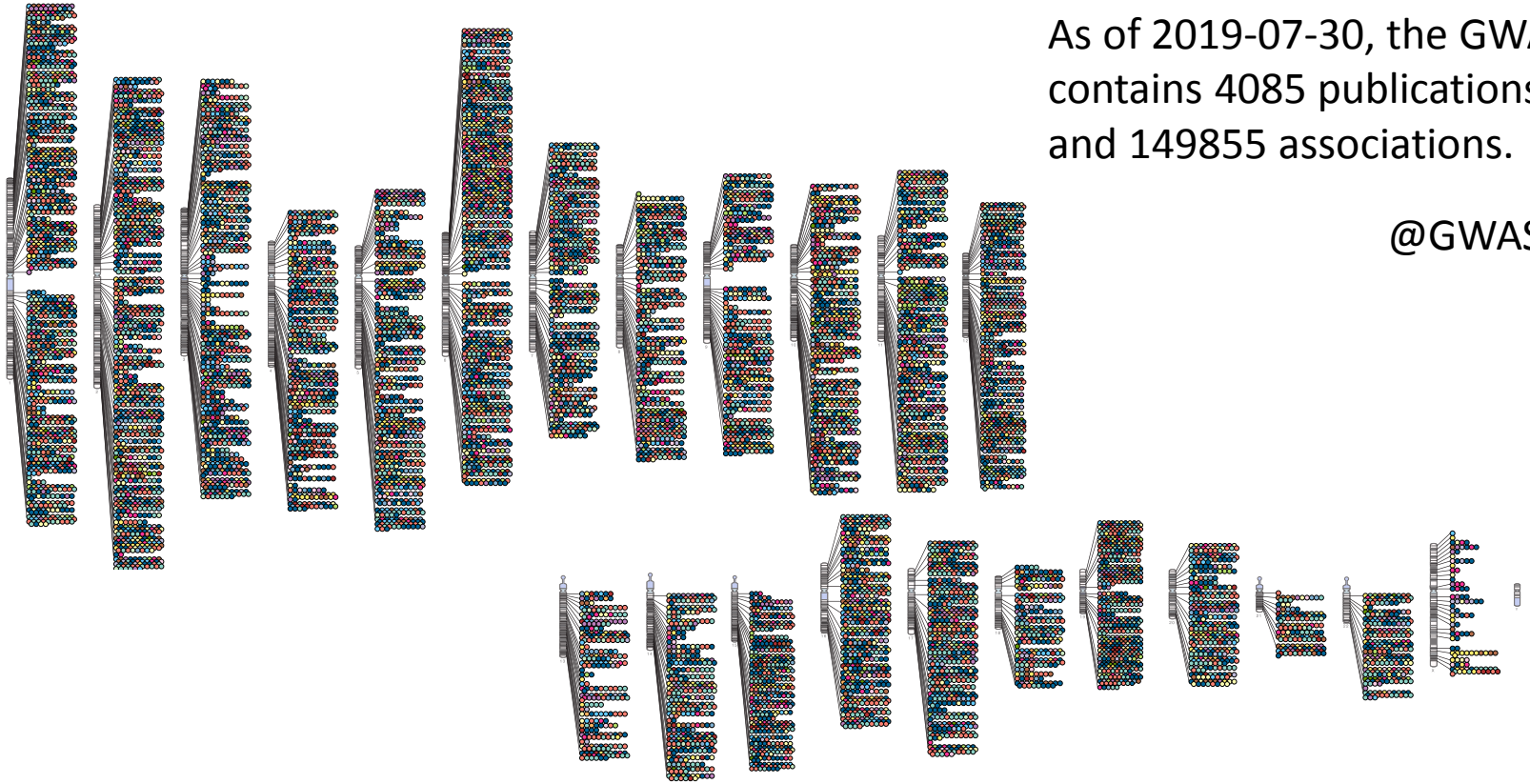
Beware the chopsticks gene!

Hamer, D., & Sirota, L. (2000). Beware the chopsticks gene.



J Novembre *et al. Nature* **000**, 1-4 (2008) doi:10.1038/nature07331





As of 2019-07-30, the GWAS Catalog contains 4085 publications and 149855 associations.

@GWASCatalog

- UK Biobank aims to improve the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses – including cancer, heart diseases, stroke, diabetes, arthritis, osteoporosis, eye disorders, depression and forms of dementia.
- It is following the health and well-being of 500,000 volunteer participants and provides health information, which does not identify them, to approved researchers in the UK and overseas, from academia and industry.

This is all nice and dandy, but how do you implement it?

This is all nice and dandy, but how do you implement it?

- Use existing software

This is all nice and dandy, but how do you implement it?

- Use existing software
- Make your own software from scratch

This is all nice and dandy, but how do you implement it?

- Use existing software
- Make your own software from scratch
- Modify existing software to work for your own needs

This is all nice and dandy, but how do you implement it?

- Use existing software
- Make your own software from scratch
- Modify existing software to work for your own needs

This is all nice and dandy, but how do you implement it?

- Use existing software
- Make your own software from scratch
- Modify existing software to work for your own needs

Then turn that into an R package 🎁 so anyone can use it! 👍

gwasurvivr: an R package for genome wide survival analysis

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Search:

Home | Install | Help | Developers | About

Home » Bioconductor 3.9 » Software Packages » gwasurvivr

gwasurvivr

platforms all | rank 1595 / 1741 | posts 0 | in Bioc 1 year
build ok | updated before release | dependencies 124

DOI: 10.18129/B9.bioc.gwasurvivr

gwasurvivr: an R package for genome wide survival analysis

Bioconductor version: Release (3.9)

gwasurvivr is a package to perform survival analysis using Cox proportional hazard models on imputed genetic data.

Author: Abbas Rizvi, Ezgi Karaesmen, Martin Morgan, Lara Sucheston-Campbell

Maintainer: Abbas Rizvi <aarizv at gmail.com>

Citation (from within R, enter `citation("gwasurvivr")`):

Rizvi A, Karaesmen E, Morgan M, Sucheston-Campbell L (2019). *gwasurvivr: gwasurvivr: an R package for genome wide survival analysis*. R package version 1.2.0, https://github.com/suchestoncampbelllab/gwasurvivr.

Documentation »

Bioconductor

- Package vignettes and manuals.
- Workflows for learning and use.
- Course and conference material.
- Videos.
- Community resources and tutorials.

R / CRAN packages and documentation

Support »

Please read the posting guide. Post questions about Bioconductor to one of the following locations:

- Support site - for questions about Bioconductor packages
- Bioc-devel mailing list - for package developers

gwasurvivr: an R package for genome wide survival analysis

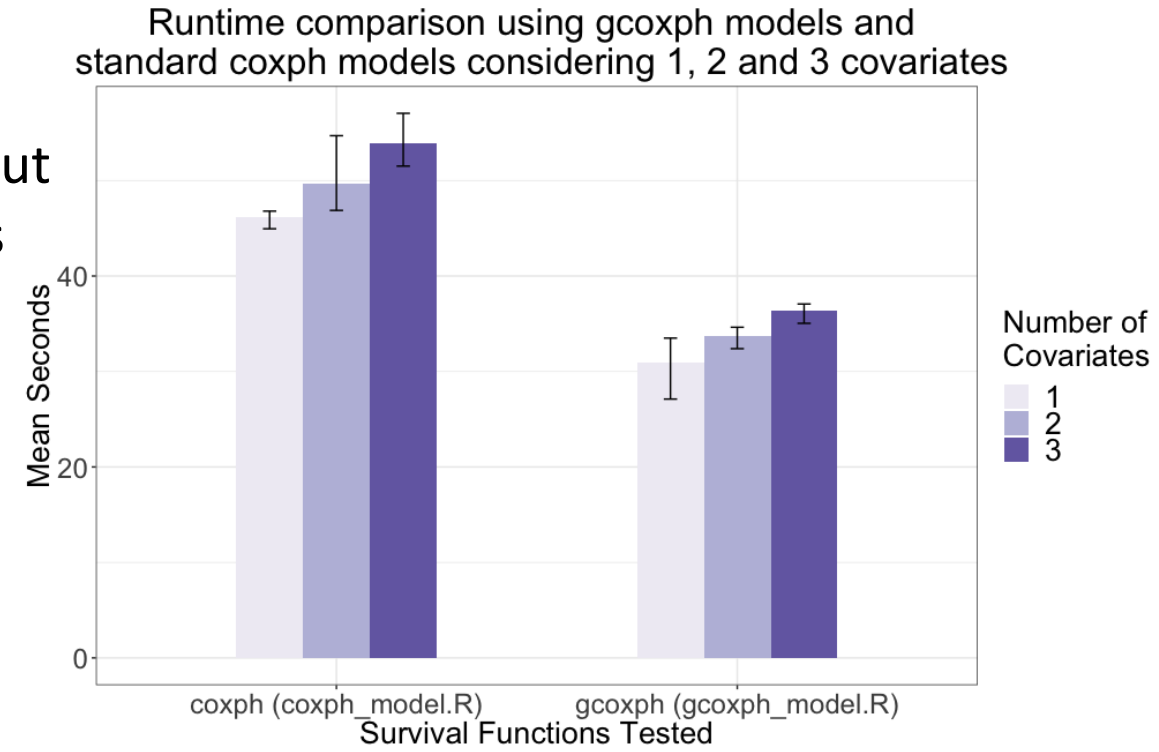
- Fast survival analysis applied genome-wide
- Can model SNP*covariate interaction
- Parallel computing across cores
- Less data wrangling for the user:
  - Filter SNP imputation/quality metrics & users can subset data by sample IDs
  - Accept output from popular imputation packages/services

Modifying classic survival package in R

Goal of modification: Potentially decrease number of iterations needed for convergence of parameter estimates

Implementation:

- Create base model without the SNP, use estimates as initial points during optimization.
- If no covariates added, parameter estimation begins at null.



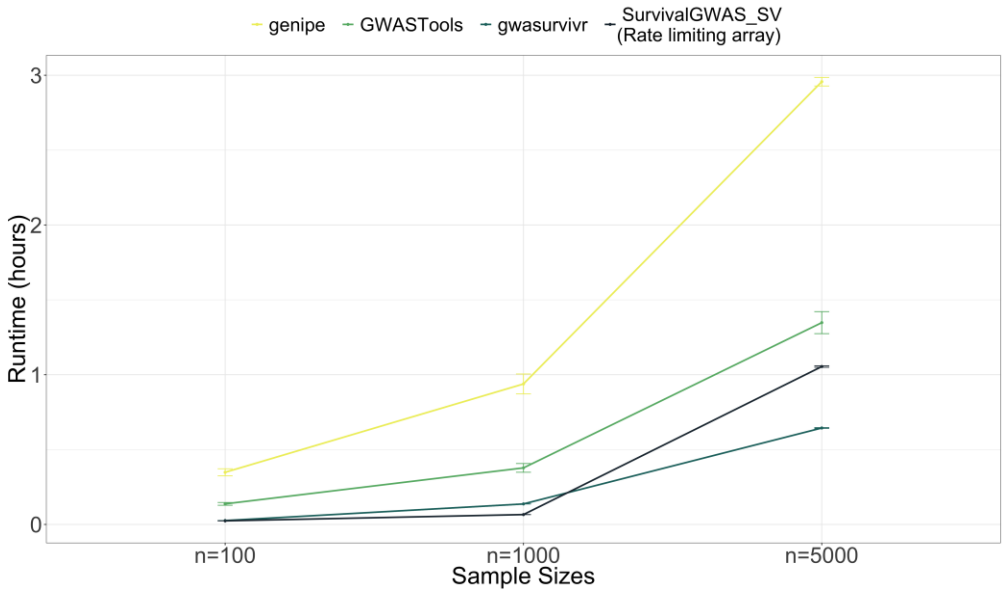
Benchmarking Simulations Design

All simulations performed using identical CPU constraints and using IMPUTE2 format

- Survival time and event (alive/dead) were simulated using a normal and binomial distribution, respectively
- Covariates were simulated using normal distributions.
- Genetic data were simulated using HAPGENv2 (Marchini et al, Nature Genetics)

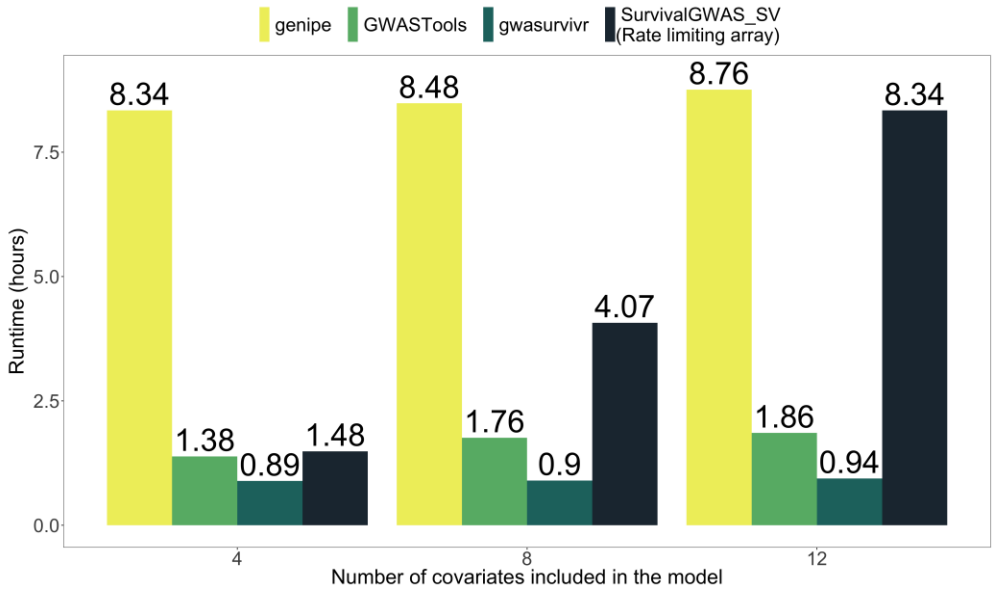
Benchmarking Experiments

- Compare run times at different Ns for 100,000 SNPs
- Compare run time if number of covariates increase
- Effect of sample size



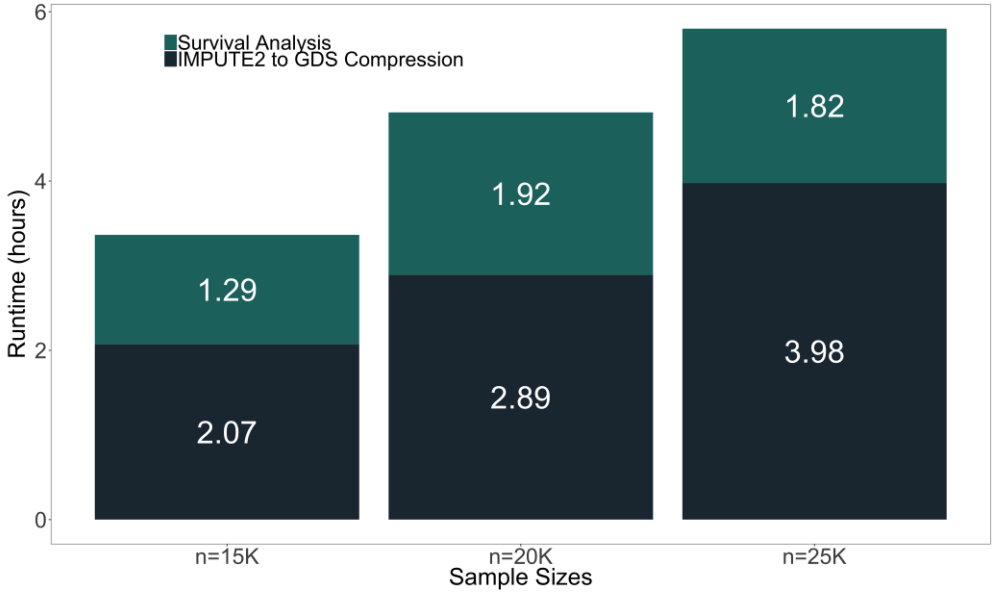
Benchmarking Experiments

- Compare run times at different Ns for 100,000 SNPs
- Compare run time if number of covariates increase
- Effect of sample size



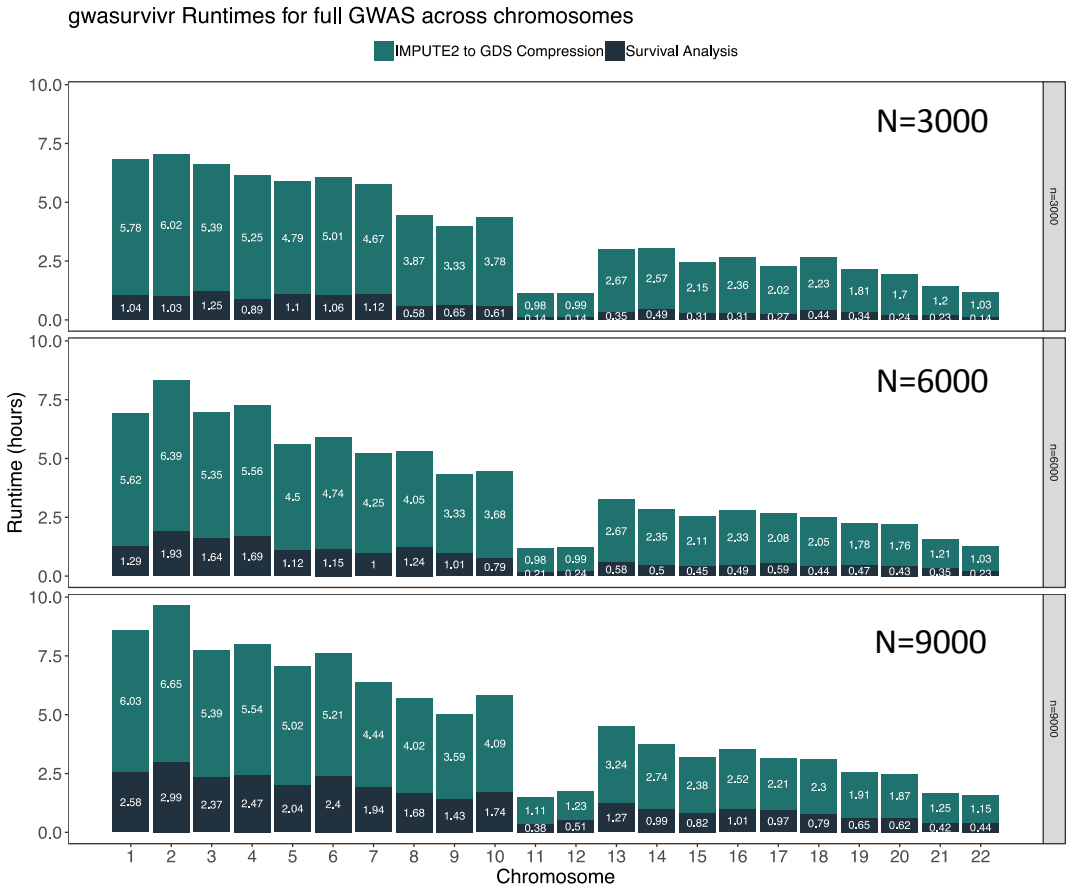
Benchmarking Experiments

- Compare run times at different Ns for 100,000 SNPs
- Compare run time if number of covariates increase
- Effect of sample size

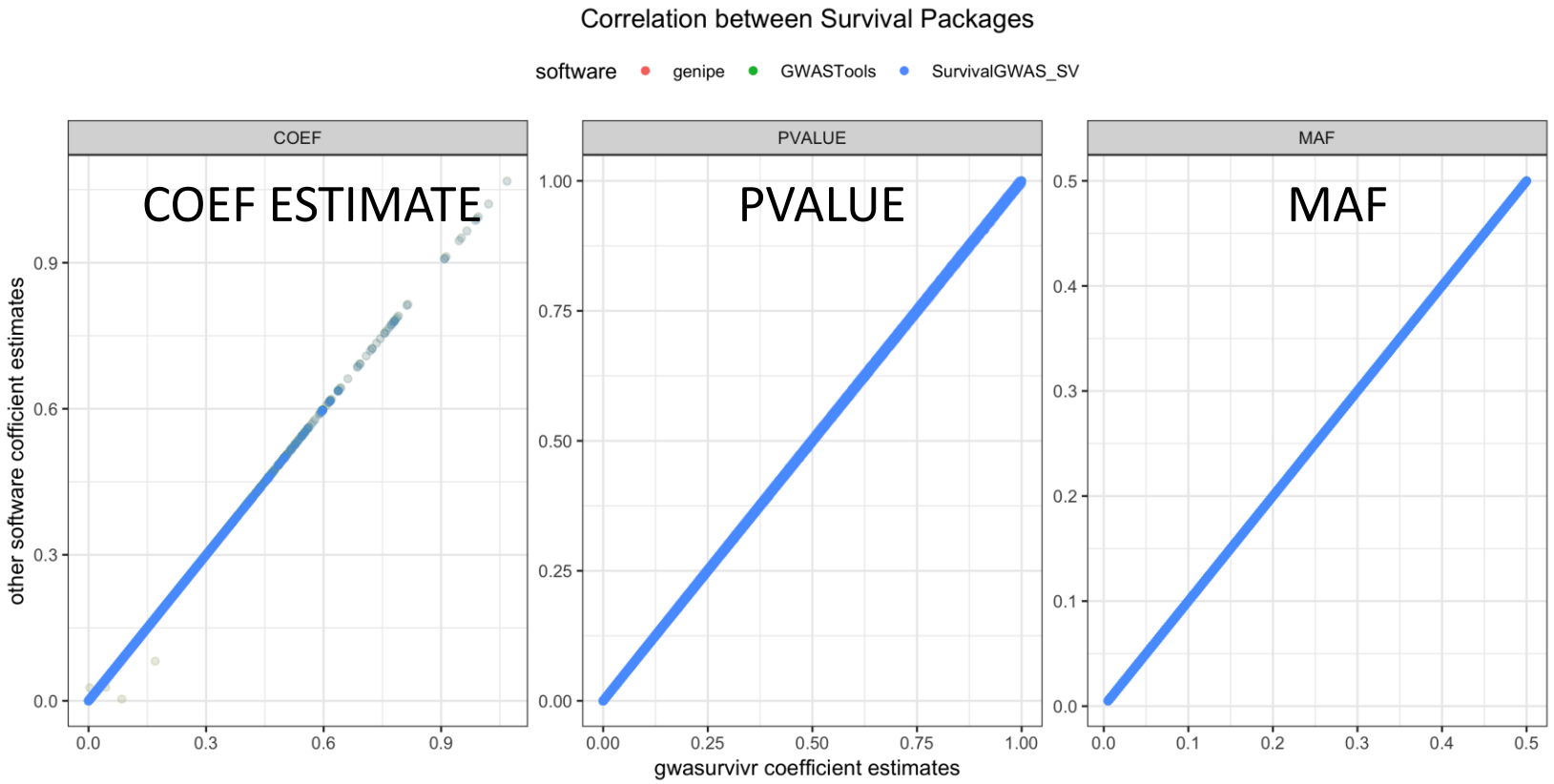


GWAS Simulations

- Simulated data from HAPGENv2 in IMPUTE2 format
- Different sample sizes of n=3000, n=6000, n=9000
- Largest GWAS still finishes in under 10 hours running in parallel



Diagnostics with other software



Conclusions

- Most of the times you don't need to reinvent the wheel, think about smart modifications to make the process more efficient.
- Don't forget to benchmark for your own sanity and start early on in the development process.
- Data wrangling before the analysis can be a challenge and big waste of time. Coming up with ways to automate it may seem like a big time sunk *in that moment*, but remember that most of the time it pays off!
- If coding in R, making packages (and you don't need to make them public!) is a great way to keep yourself sane and facilitate reproducibility.

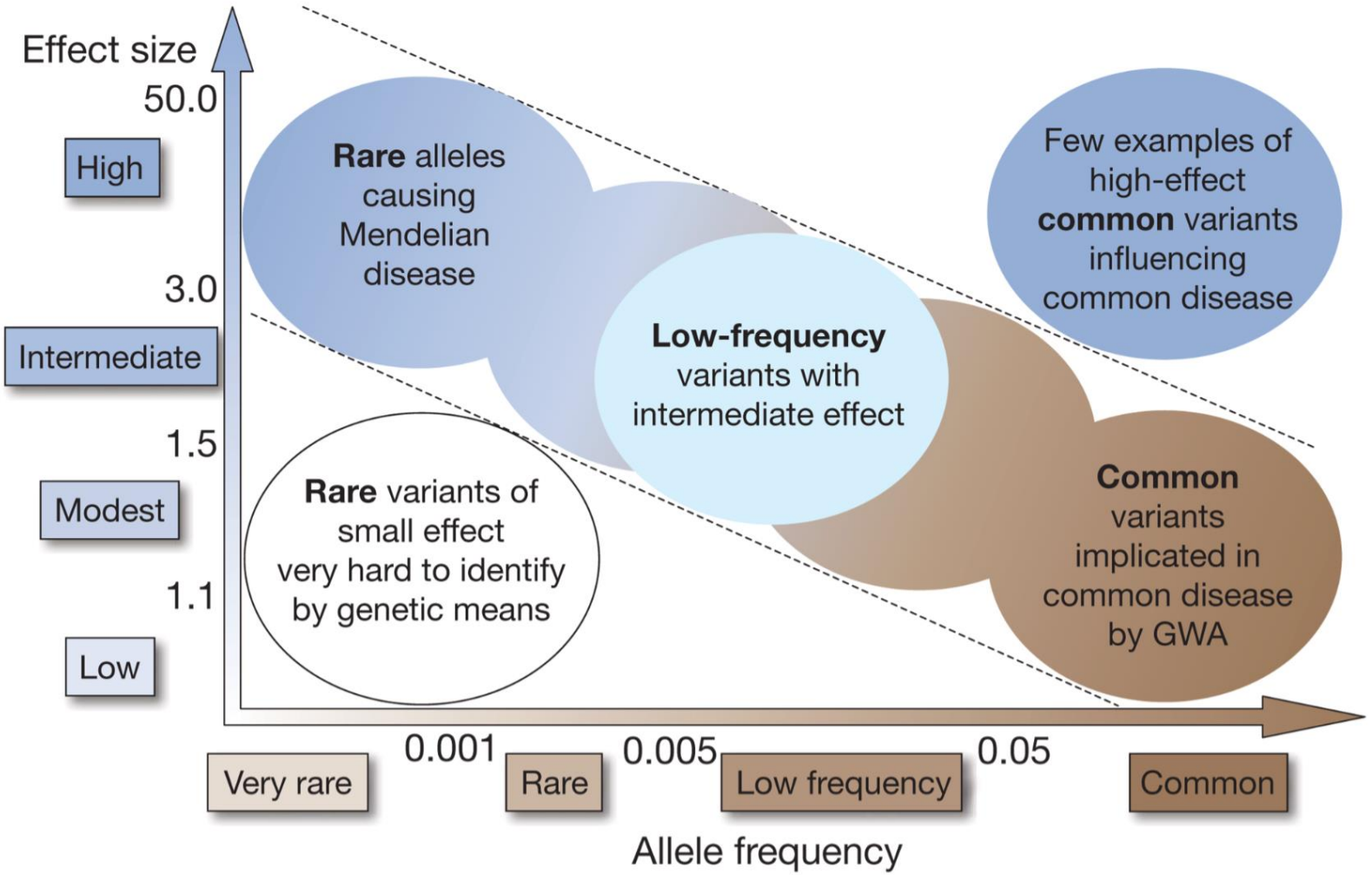
Thanks for listening!

If you are looking for a R meetup don't forget to check out R Ladies

meetup.com/RLadies-Columbus

Twitter @RLadiesColumbus

Missing heritability concept



Manolio, Teri A., et al. *Nature* 461.7265 (2009): 747-753.