

# The Role of AI

Algorithm to bedside symposia

>>>





# DATA STORIES

## From algorithm to bedside

28-01-2025

@NerminGhith



Nermin Ghith, dual B.Sc. in bio/chem.; MPH, PhD  
Senior Forsker  
Gentofte or Herlev Hospital  
03-14 Feb. 2025

# Octopus- Senior Epidemiologist

## Projects

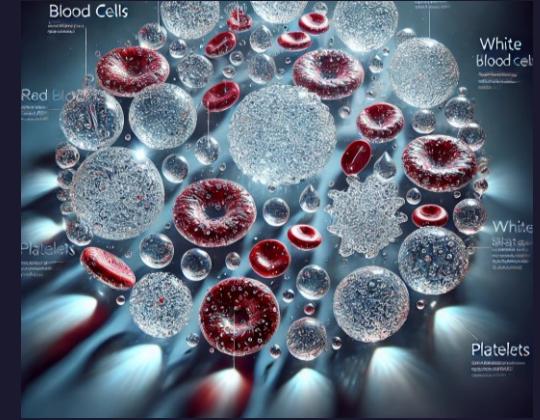
Combine and integrate large-scale data:

- biological samples and data, contextual information (family, institutional, environmental and community-specific effects), risk factors and lifestyle, clinical interventions, treatments, and health outcomes

Model and understand

- The determinants of health and causes of illness in the human population

Nesting a coherent scientific insight to advance clinical care, health policies, and population health!



# Data

- Danish and Nordic registries
- WHO national, regional, and global databases, surveys, and questionnaire data
- Industry datasets (e.g, GIDEON, IQVIA)
- Own projects' data (e.g., interviews with patients and healthcare professionals).

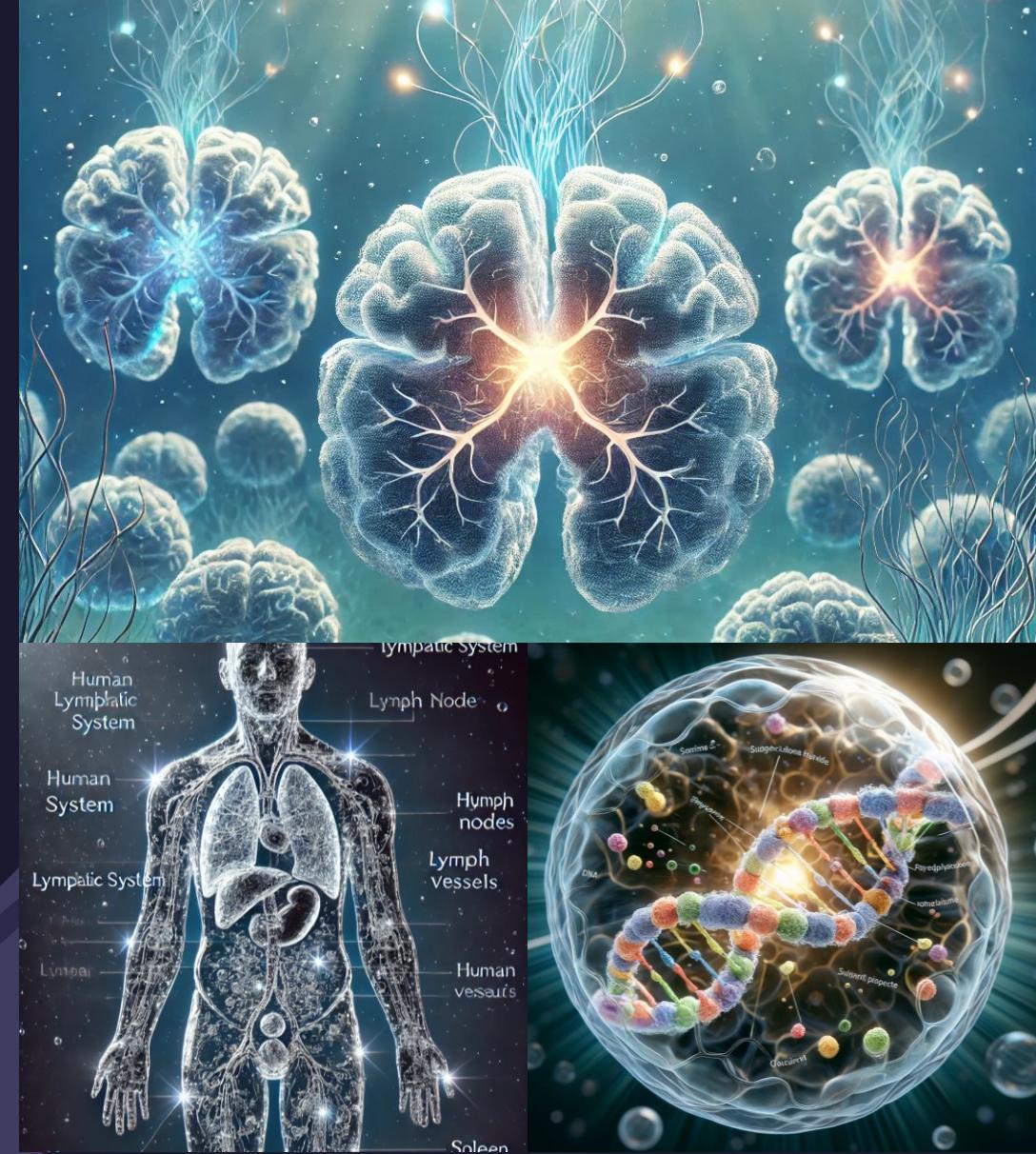


# My research and technical areas

Register-based Epidemiology / Health Data Science

# (Convergence of Epidemiology, Biostatistics, and Data Science)

- Late effects, drug use and QoL (SeniorForsker/PostDoc,WHO)
  - (MetaGenomics/Microbiome): Global Burden of Antimicrobial resistance and infectious diseases (DTU,WHO)
  - Multimorbidity and polypharmacy (seniorForsker/PostDoc,WHO)
  - Global and Regional Burden of Disease (DTU,WHO)
  - Contextual epidemiology, Clinical care, and health outcomes (Ph.D, WHO)
  - Behavioral Risk Factors (e.g., smoking, substance use, diet ) – WHO, master thesis



# Lunch symposia

Omics-Based Clinical and Population Studies (session 1:  
Mon. 3 Feb. 2025 - Specific Case Studies)

Contextual Epidemiology of Cardiometabolic  
Conditions and Multi-morbidity (session 2: Thurs. 6  
Feb. 2025 )

Scientific Methods in Biomedical Research (sessions 3  
and 4: Mon and Thurs. 10 and 13 Feb. 2025)

- Omics in biomedical research
- Global Burden of Disease, Health Analytics, and Access to Therapeutics
- Evidence and analytics



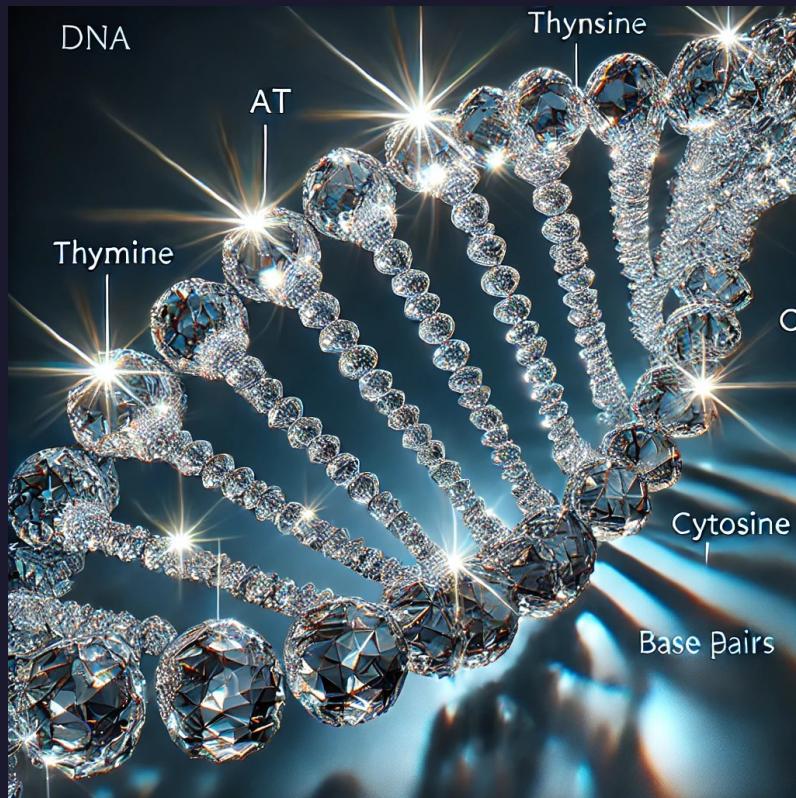
# Algorithm to bedside Omics-Based Clinical and Population Studies (session I-I)



# Antimicrobial Resistance : Global burden of AMR – Microbiome and Metagenomics



# Background



## AMR

Inequality in global health outcomes such as antimicrobial resistance (AMR) is shaped by macro determinants of health within and between countries.

## Aim

This project aims to assess the potential relationship between the community (at the city level) progress towards achieving the sustainability development goals (SDG) and the burden of antimicrobial resistance (AMR).

## Framework

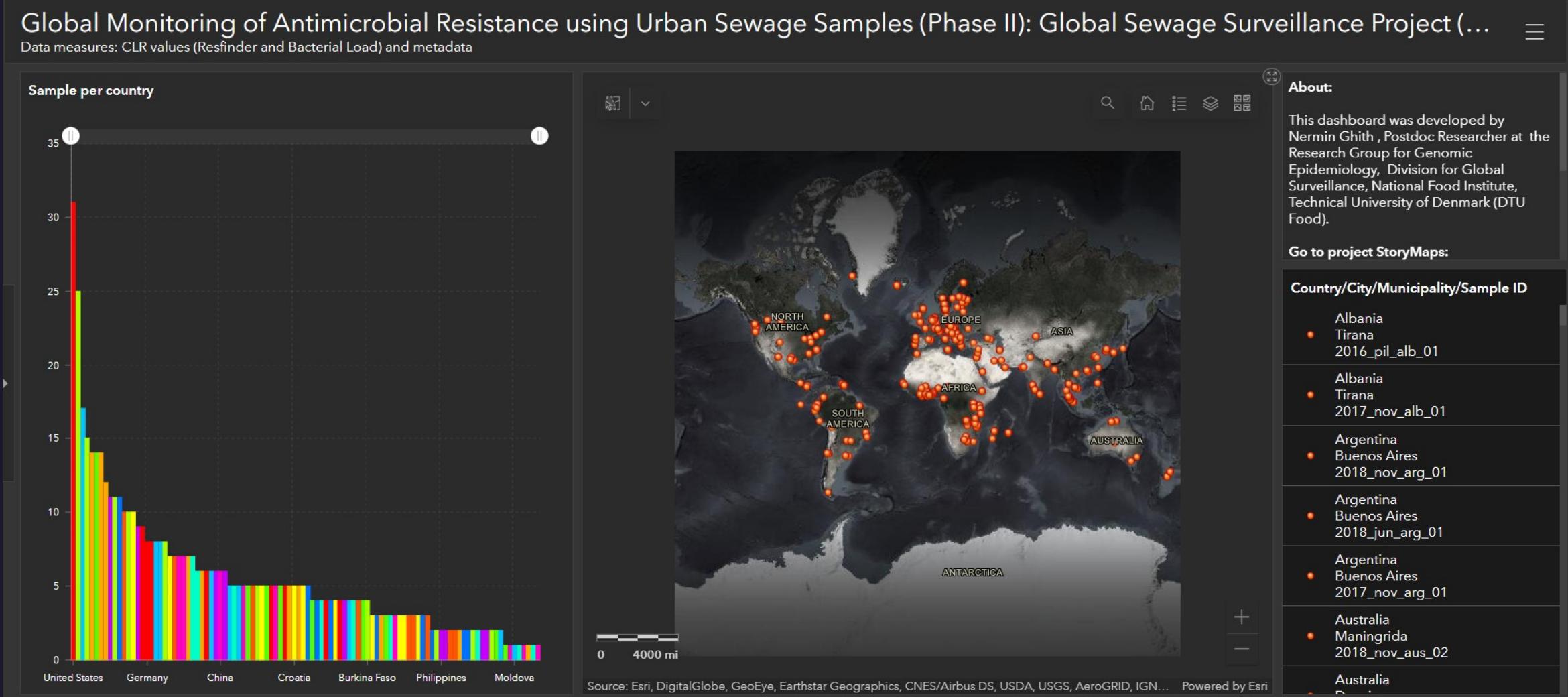
The study identifies classes of antimicrobial resistance that varies between the sampling sites within cities, and compliance with SDGs goals that might explain variation in the global burden of AMR between countries.

# DTU research role (using samples from 104 countries and more than 230 cities)

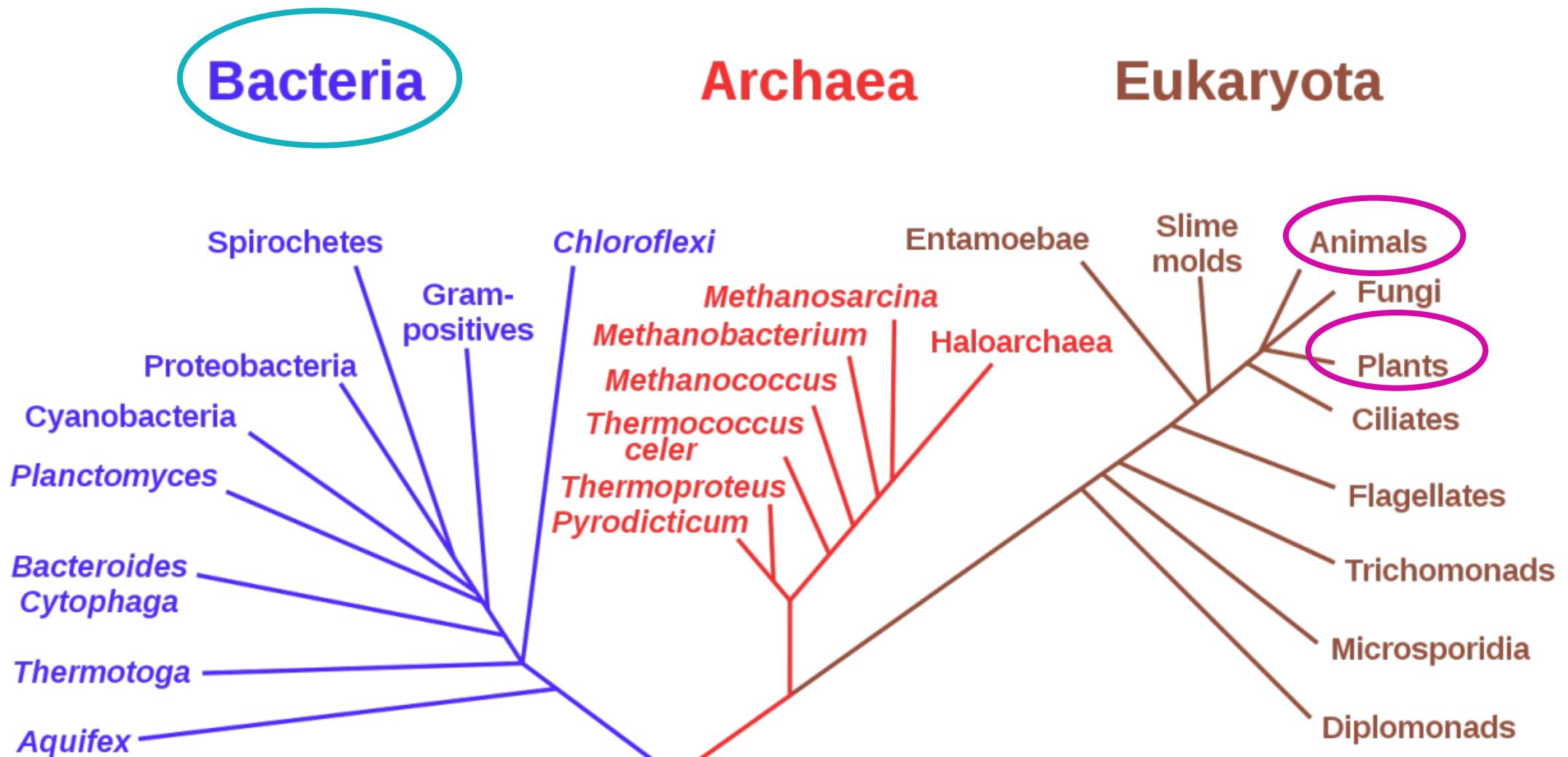
- Identify **macro** determinants of AMR
- Assess the dual burden of AMR and **infectious** diseases
- Assess the relationship between **morbidity** and **mortality** profiles (i.e., childhood and adult cancer; and cardiometabolic conditions) and the global burden of AMR
- Assess the relationship between **SDGs** and AMR



## Own data



# What are microorganisms?



Potential good source for the tree of life:

<https://biosciences.gatech.edu/research/faculty-research-video-gallery>

## The Oral and Gut Bacterial Microbiomes: Similarities, Differences, and Connections

- July 2020
- Biological Research for Nursing 23(1):109980042094160
- DOI:
- [10.1177/1099800420941606](https://doi.org/10.1177/1099800420941606)

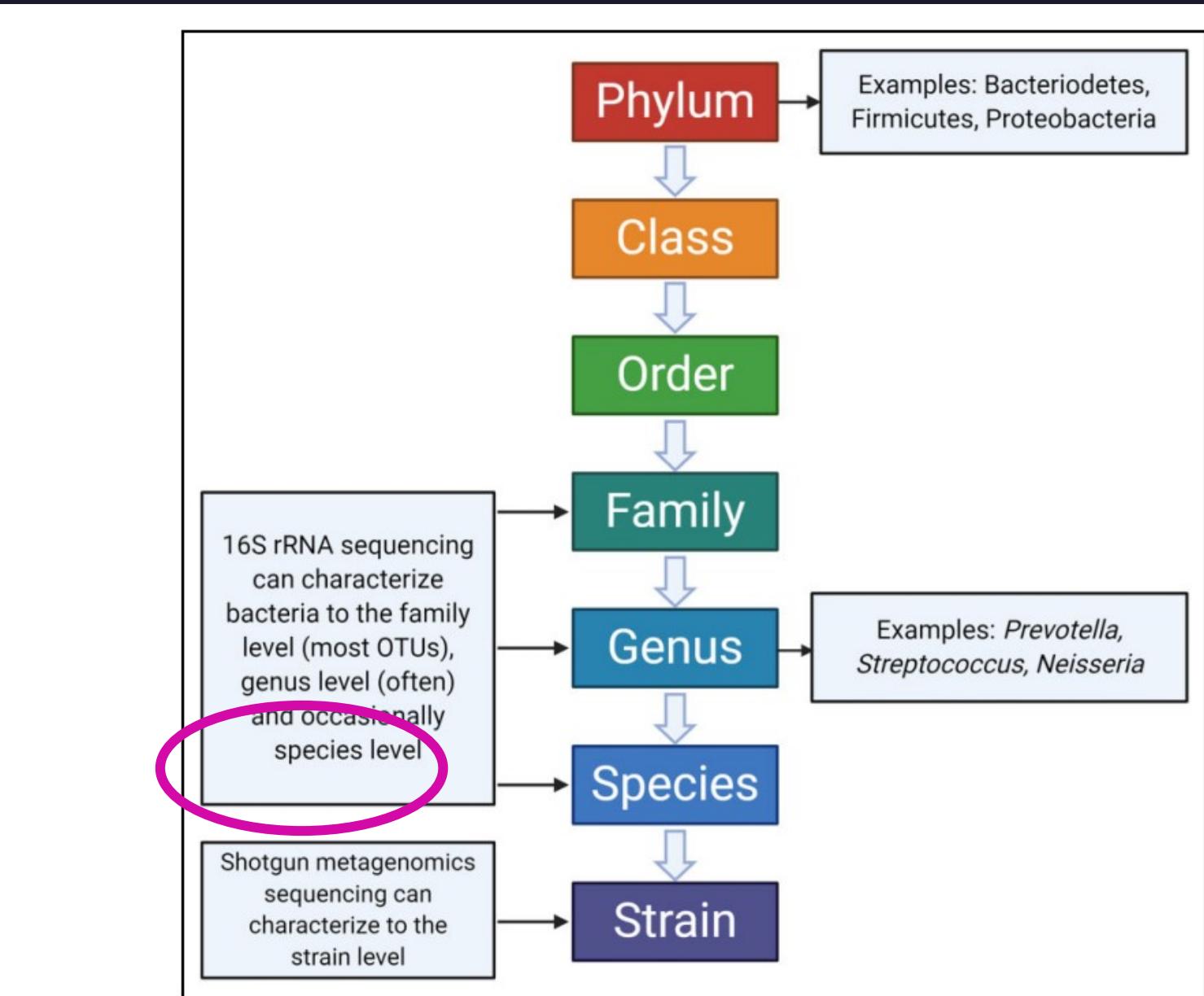
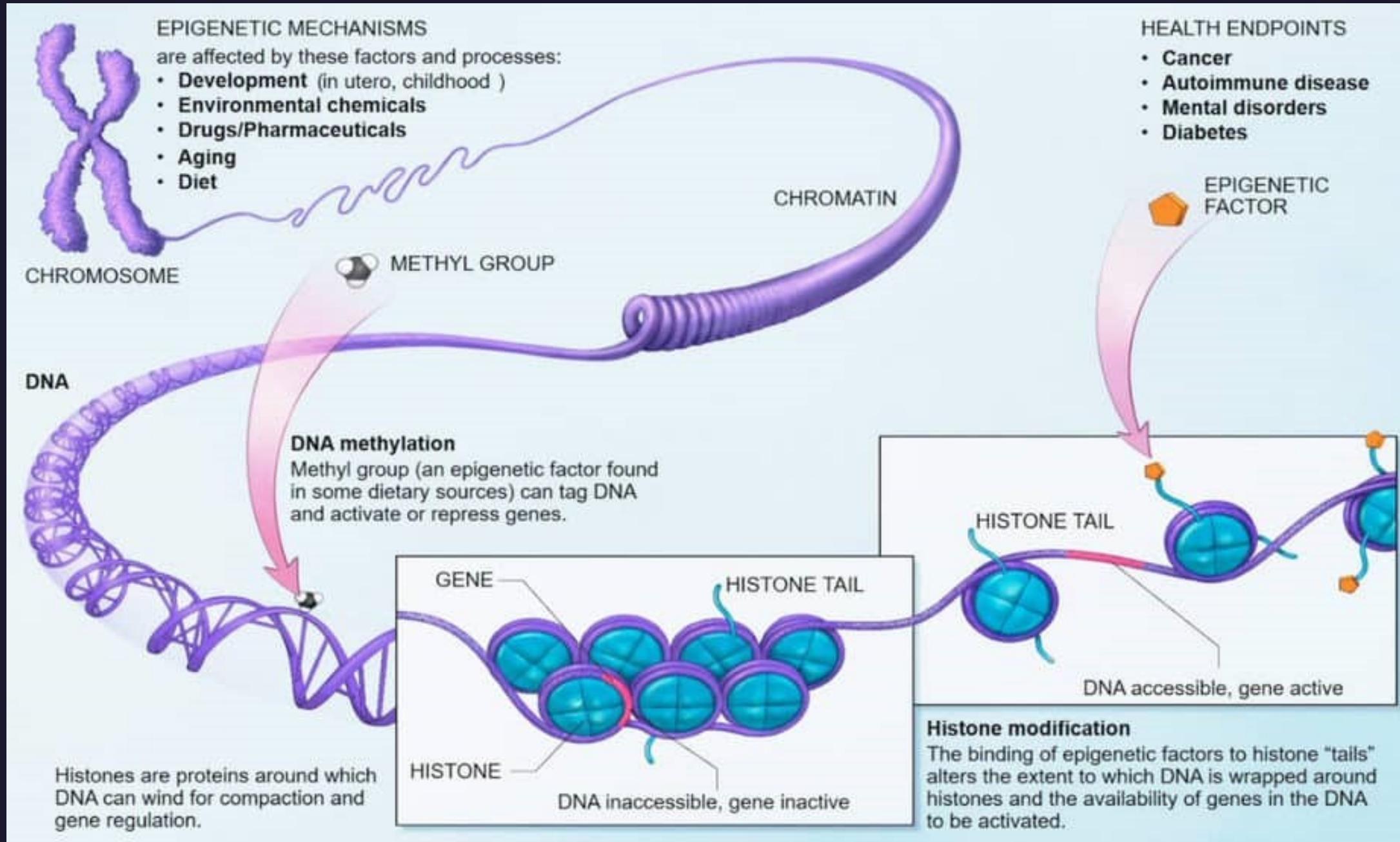
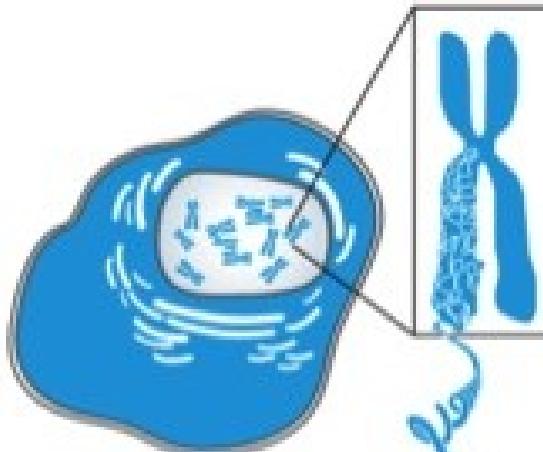


Figure 5. Taxonomic classification of bacteria from phylum to strain level.

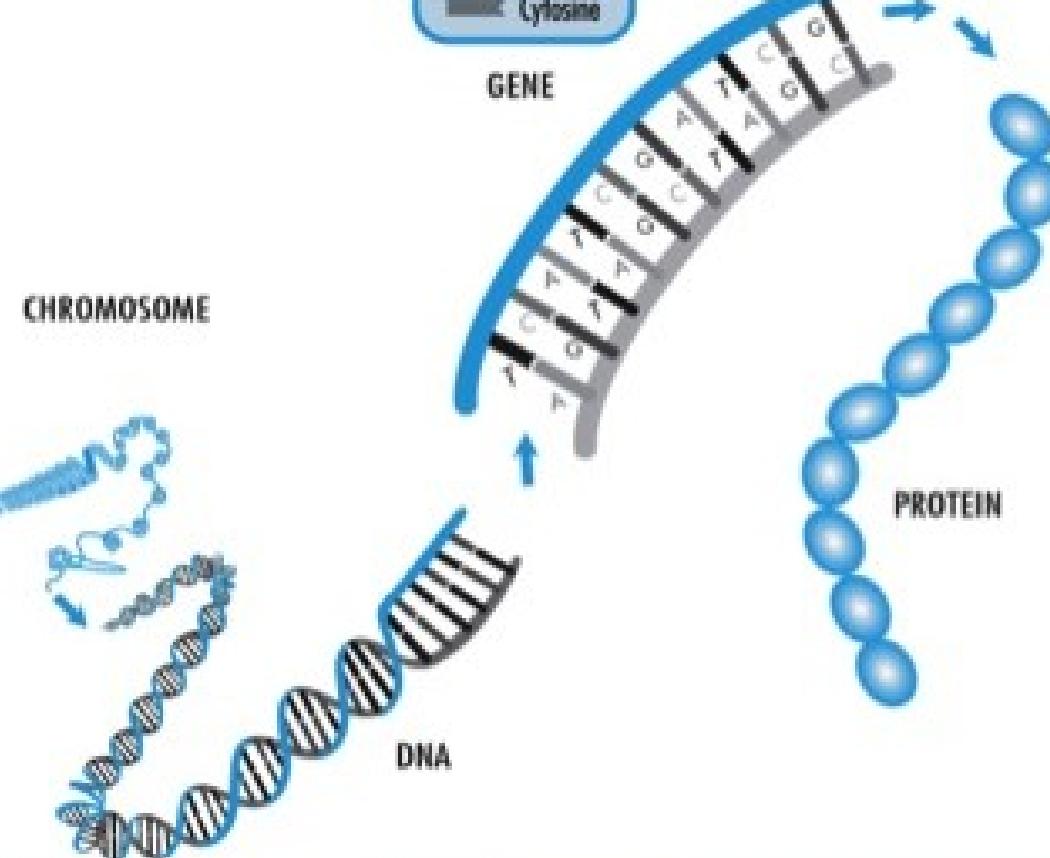


## Chromosome to Gene to Protein



CHROMOSOME

CELL



DNA

PROTEIN



CHROMOSOME



DNA

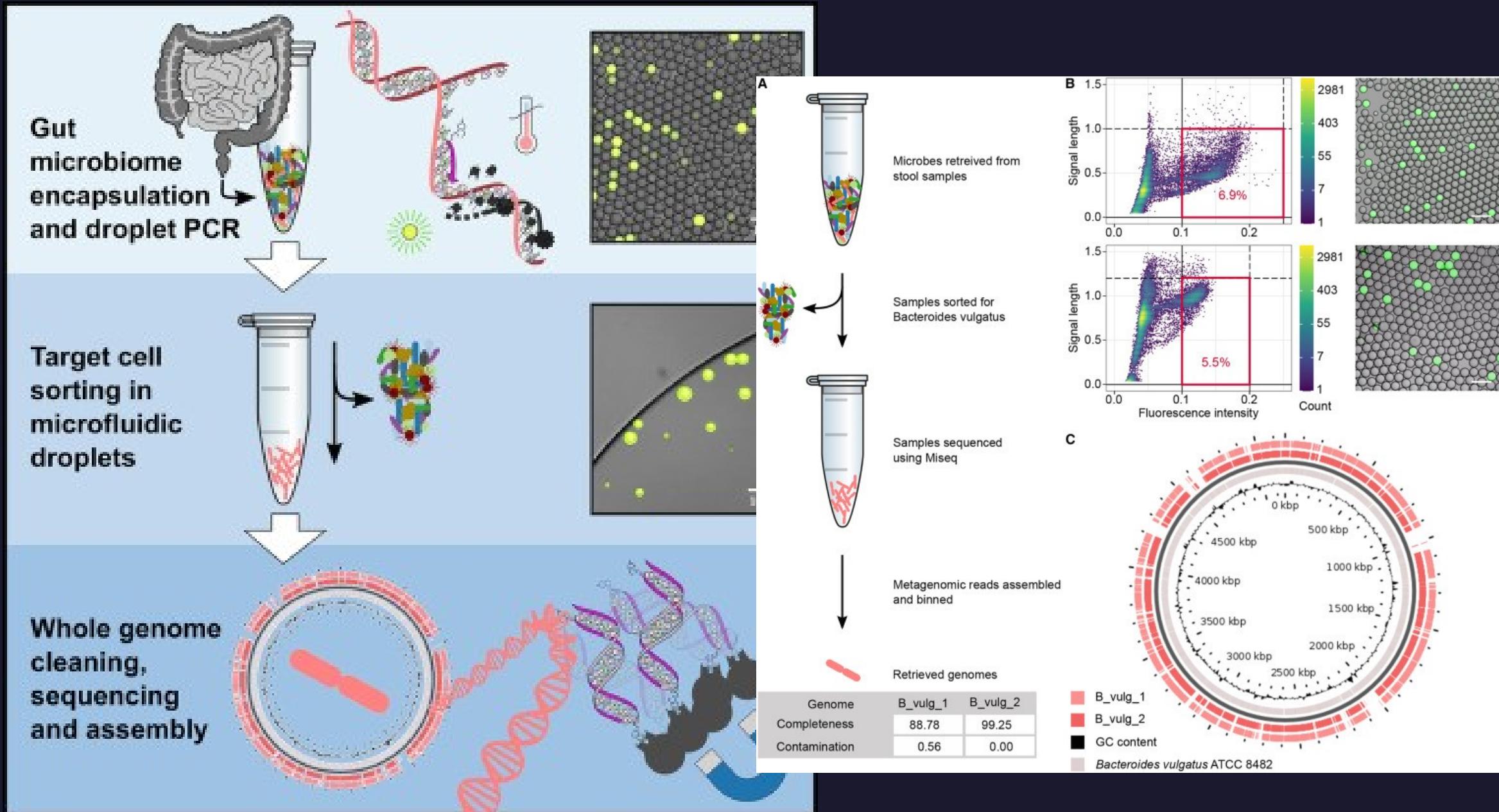


GENE

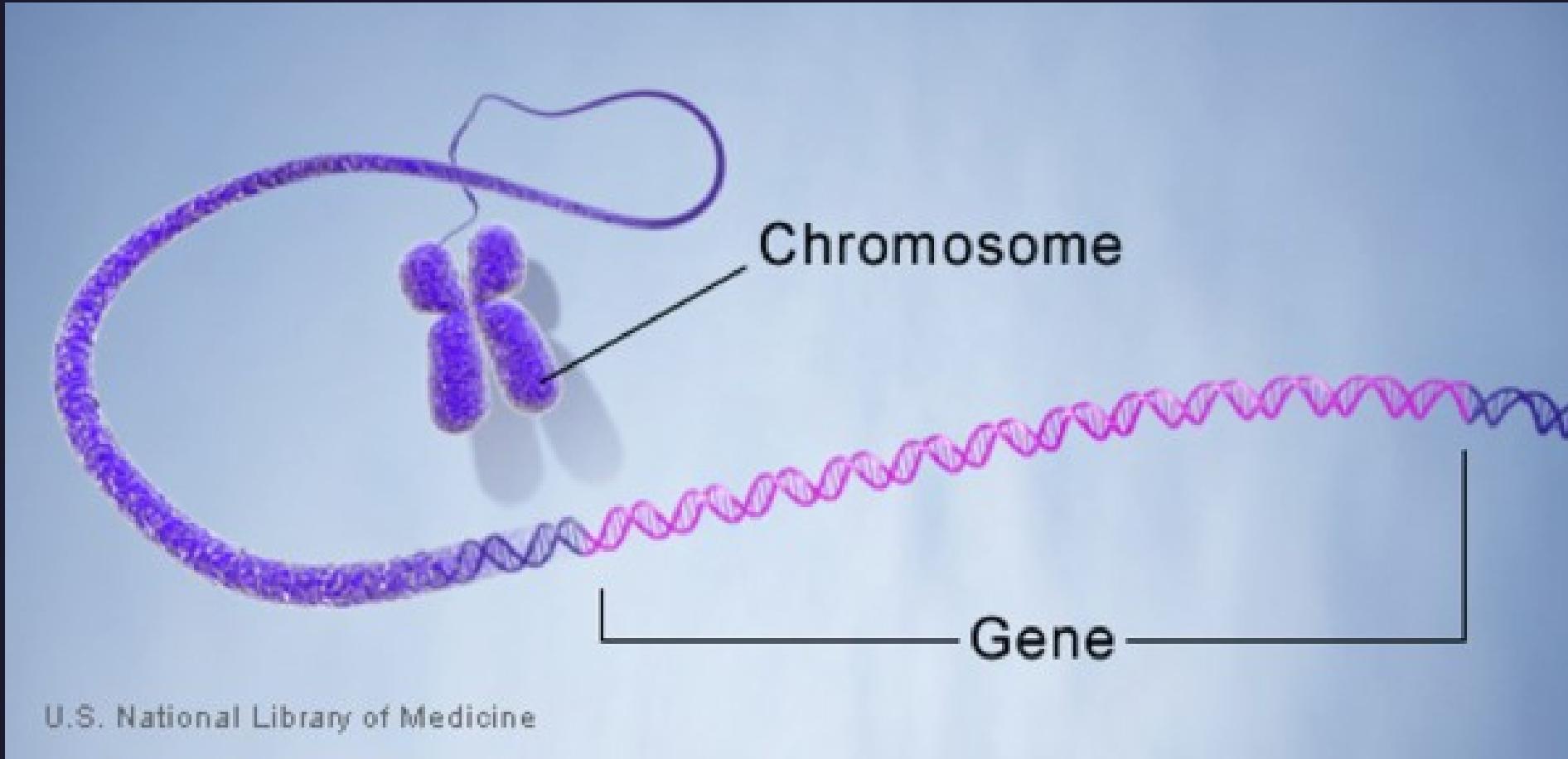


PROTEIN

Source: <http://epilepsygenetics.net/2016/05/25/explaining-variants-of-uncertain-significance-a-guide-for-clinicians/>



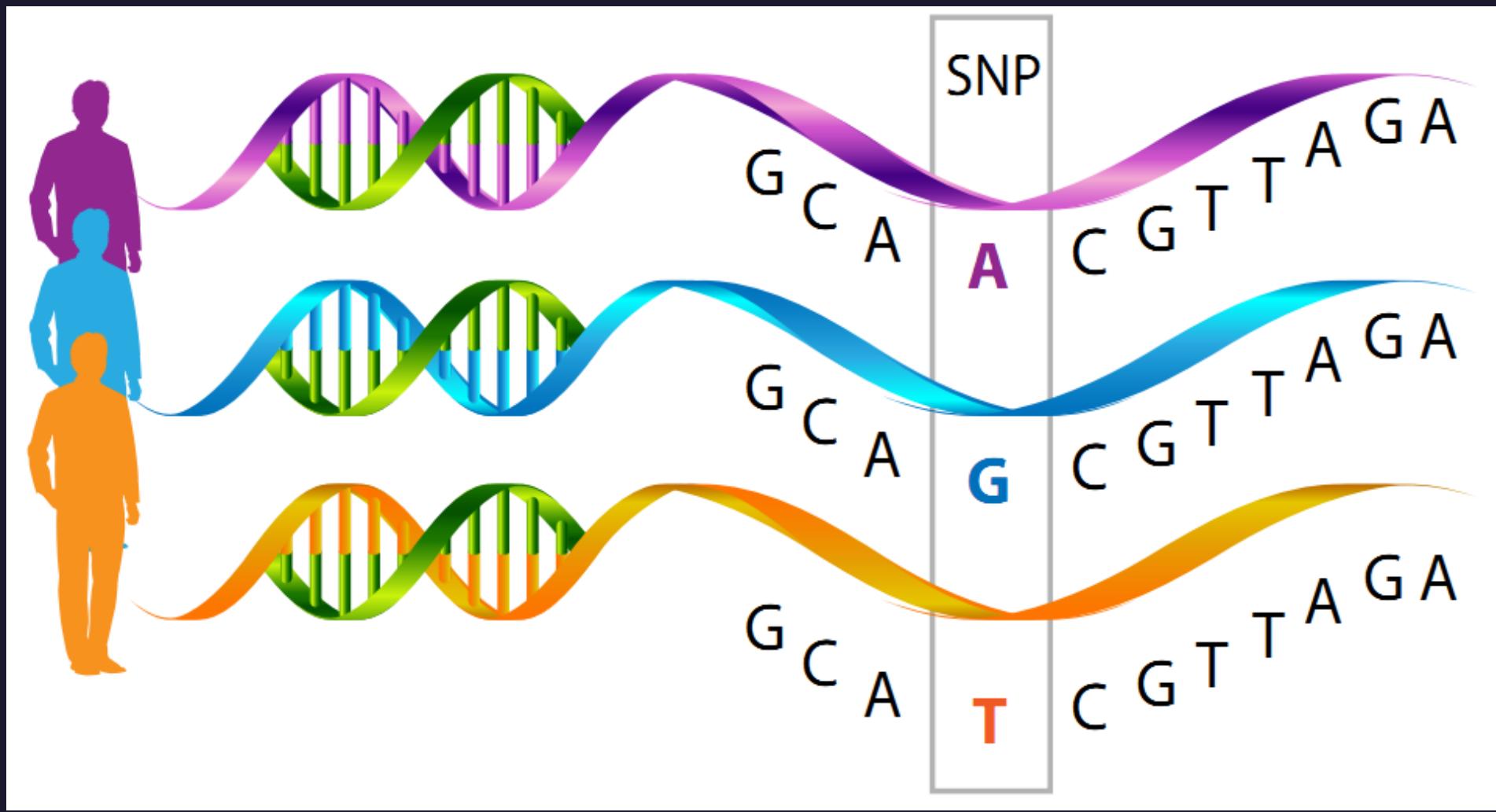
Source: <https://doi.org/10.1016/j.crmeth.2021.100137>



U.S. National Library of Medicine

A gene is the basic physical and functional unit of heredity. Genes are made up of DNA. Some genes act as instructions to make molecules called proteins, which are needed for the body to function. However, many genes do not code for proteins, instead they help control other genes.

Source: <https://medlineplus.gov/genetics/understanding/basics/gene/>



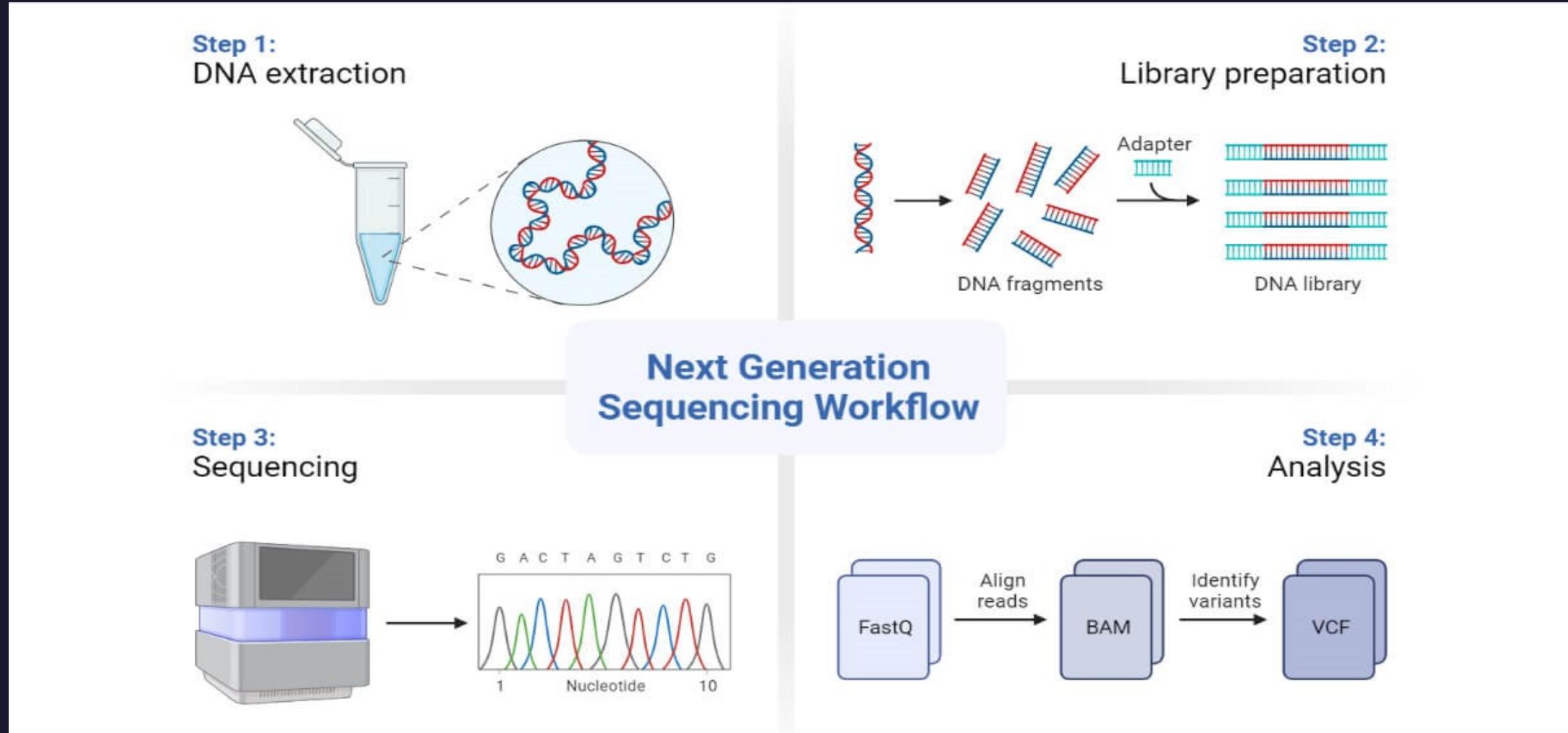
SNPs, or Single Nucleotide Polymorphisms (pronounced snips), are single base changes within genes, akin to single letter spelling variations. (Think analyse vs analyze.) It is important to emphasize that these are variations, NOT mutations. SNPs represent 90+% of the total variation amongst individuals, meaning that nearly every genetic difference you encounter on a daily basis is caused by single nucleotide changes in various genes. These tiny variations (SNPs) may slightly change the function of the protein encoded by the corresponding gene.

Source: <https://www.nutrigeneticspecialists.com/single-post/2017/03/27/what-is-a-snp>

28-01-2025

@NerminGhith

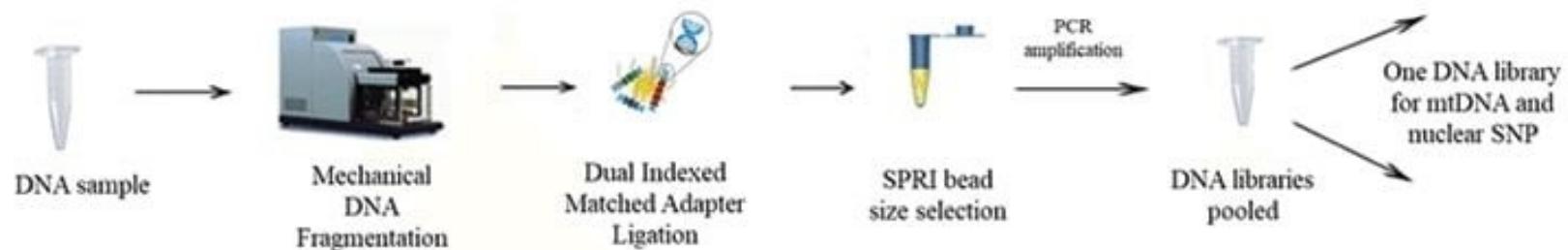
18



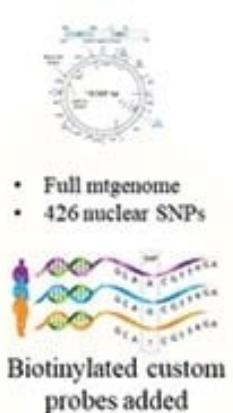
**NGS** is a powerful technology that has transformed the health data landscape. Among many other benefits, it has drastically improved our ability to determine the presence of AMR genes (bacterial genes known to confer resistance to an antimicrobial drug)

Source: <https://microbenotes.com/next-generation-sequencing/ngs/>

## 1) DNA Fragmentation and Library Preparation



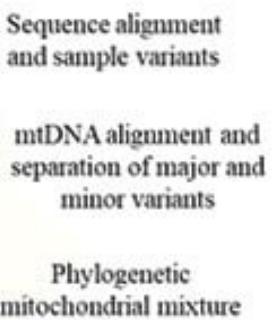
## 2) Probe Capture and Enrichment



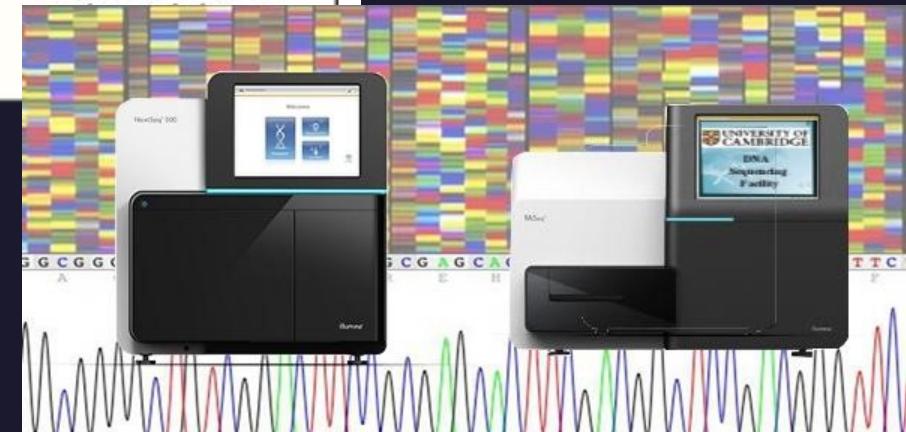
## 3) Next Gen Sequencing



## 4) Data Analysis



<https://doi.org/10.3390/genes9010049>



Source: University of Cambridge, Research Facilities at the Department of Biochemistry  
<https://facilities.bioc.cam.ac.uk/dna-sequencing-overview/next-generation-sequencing-ngs>

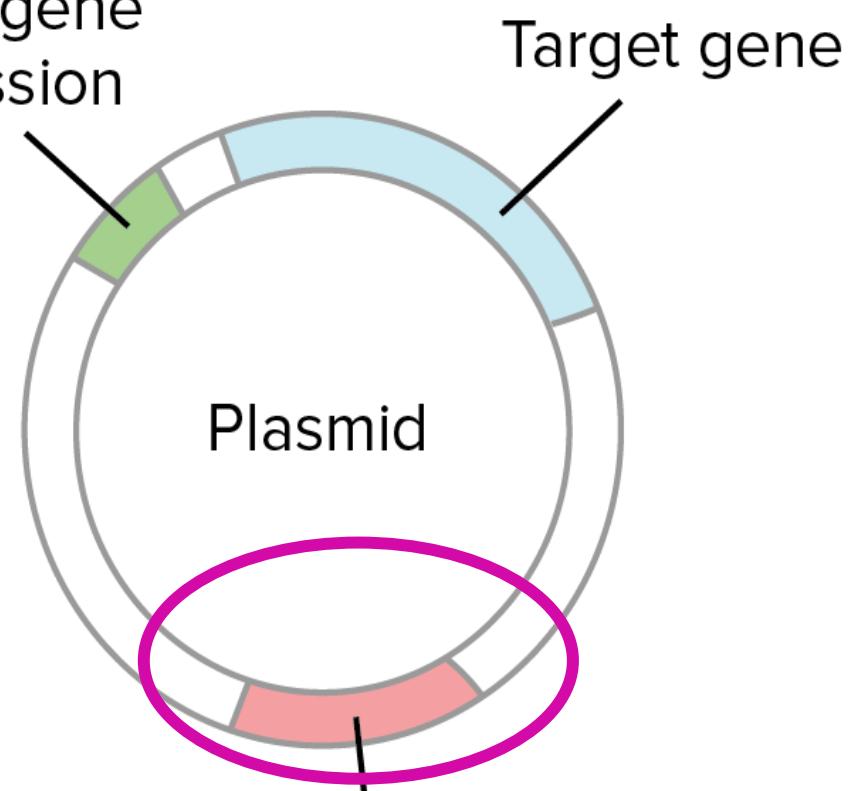
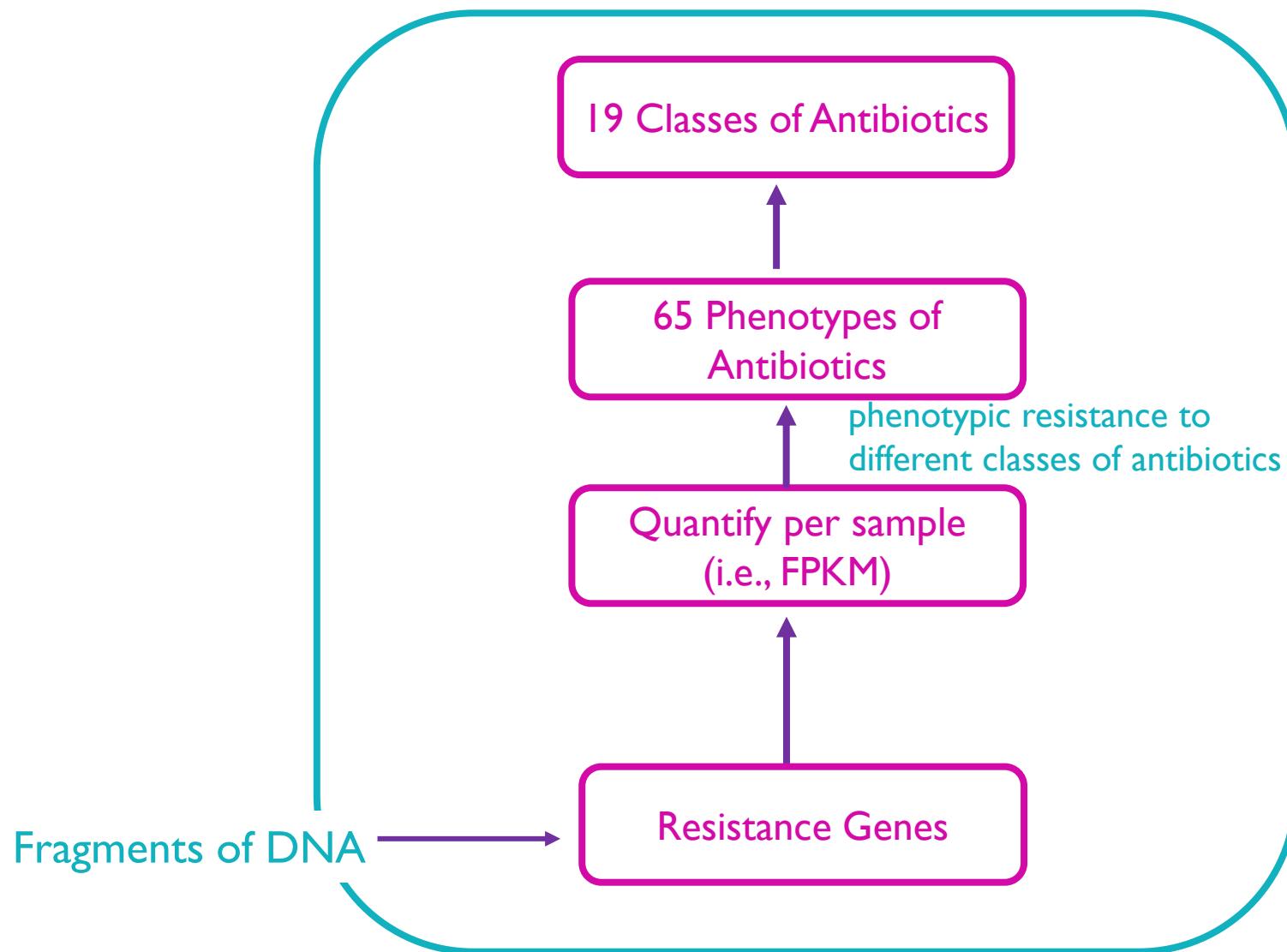
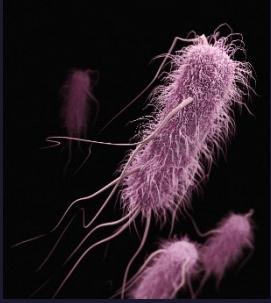


Figure source link: <https://www.khanacademy.org/science/biology/biotech-dna-technology/dna-cloning-tutorial/a/bacterial-transformation-selection>

## Own Project Pipeline



Whole-genome sequencing output  
provides a rough estimation of  
bacteria species abundance



## Bacterial species

A	B	C	D	E	F	G	H	I	J	K
species_name	genomic_2									
[Bacillus] caldolyticus	0	0	0	0	0	0	0	0	0	0
[Bacillus] clarkii	0	0	0	0	0	0	0	0	0	0
[Bacteroides] pectinophilus	3719	419	4260	1882	4858	609	49	613	388	327
[Brevibacterium] flavum	0	0	0	0	0	0	0	0	0	0
[Clostridium] aerotolerans	0	0	0	0	1	0	0	0	0	0
[Clostridium] aminophilum	0	0	0	0	1	1	0	0	1	0
[Clostridium] asparagiforme	396	0	260	130	204	101	1	115	53	98
[Clostridium] bolteae	747	42	702	59	599	160	31	98	67	590
[Clostridium] cellulosi	0	0	0	0	0	0	0	0	0	0
[Clostridium] citroniae	99	1	110	52	131	6	4	6	3	203
[Clostridium] clostridiiform	320	52	246	69	323	180	11	194	58	211
[Clostridium] coelatum	0	1	3	0	3	0	1	0	1	0
[Clostridium] dakarens	54	4	4	4	6	2	1	1	4	4
[Clostridium] fimicarium	73	1	8	3	5	5	1	4	2	1
[Clostridium] glycyrrhizinilyt	851	153	647	275	588	248	57	206	158	103
[Clostridium] hiranonis	8	0	3	0	2	0	0	1	1	0
[Clostridium] hylemonae	0	0	2	1	0	0	1	7	1	3
[Clostridium] innocuum	0	10	8	0	0	34	0	35	17	0
[Clostridium] leptum	470	66	308	82	320	547	88	511	144	70
[Clostridium] methylpentos	108	4	68	1	94	50	0	1	4	5
[Clostridium] paradoxum	11	3	10	3	9	0	0	3	2	7
[Clostridium] polysaccharoly	0	0	0	0	0	0	0	0	0	0
[Clostridium] populeti	1	5	10	2	15	2	0	3	1	3
[Clostridium] perfringens	304	64	121	24	141	125	45	100	64	42
species.count.mat										

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
new_complete	country	city	latitude_in	longitude_country	Region	income_gr	Income	Country_C	SDG_Index	SDG_Index	SDG_Index	Population	Goal1Score	Goal2Score	Goal3Score	Goal4Score	Goal5Score	Goal6Score	Goal7Score
DTU_2016_1	Albania	Tirana	41.32	19.9	ALB	Europe & UpperMidc	3 ALB	71.02186	highMidd	71.40243	2877800	91.6375	60.00486	81.09271	95.155	53.57425	75.9076	75.9076	75.9076
DTU_2016_4	Australia	Woden	-35.35	149.09	AUS	East Asia & High	4 AUS	75.57571	high	77.20763	25499881	99.61	58.74013	95.58357	99.40538	81.00025	94.9532	94.9532	94.9532
DTU_2016_9	Australia	Melbourne	-37.81	144.96	AUS	East Asia & High	4 AUS	75.57571	high	77.20763	25499881	99.61	58.74013	95.58357	99.40538	81.00025	94.9532	94.9532	94.9532
DTU_2016_14	Austria	Vienna	48.21	16.37	AUT	Europe & High	4 AUT	82.08009	high	77.20763	9006400	99.467	73.76713	92.10831	99.20504	83.08375	89.5994	89.5994	89.5994
DTU_2016_18	Bulgaria	Sofia	42.7	23.32	BGR	Europe & UpperMidc	3 BGR	73.18489	highMidd	71.40243	6948445	97.507	66.34725	79.553	67.553	72.472	70.8148	70.8148	70.8148
DTU_2016_22	Brazil	Belem	-1.45	-48.47	BRA	Latin Amer UpperMidc	3 BRA	71.33911	highMiddle	68.57615	2.13E+08	86.9855	69.89925	77.83657	87.65767	70.29175	86.2456	86.2456	86.2456
DTU_2016_25	Brazil	Belem	-1.45	-48.47	BRA	Latin Amer UpperMidc	3 BRA	71.33911	highMiddle	68.57615	2.13E+08	86.9855	69.89925	77.83657	87.65767	70.29175	86.2456	86.2456	86.2456
DTU_2016_28	Botswana	Palapye	-22.54	27.12	BWA	Sub-Saharan UpperMidc	3 BWA	61.92338	lowMidd	51.92682	2351625	54.784	41.41138	53.73014	59.67133	70.16675	70.5678	70.5678	70.5678
DTU_2016_32	Canada	Regina	50.45	-104.62	CAN	North Amer High	4 CAN	79.1554	high	77.20763	37742157	99.5335	67.3625	93.95292	99.58704	82.39025	85.8998	85.8998	85.8998
DTU_2016_35	Canada	Calgary	51.05	-114.06	CAN	North Amer High	4 CAN	79.1554	high	77.20763	37742157	99.5335	67.3625	93.95292	99.58704	82.39025	85.8998	85.8998	85.8998
DTU_2016_40	Canada	Toronto	43.65	-79.39	CAN	North Amer High	4 CAN	79.1554	high	77.20763	37742157	99.5335	67.3625	93.95292	99.58704	82.39025	85.8998	85.8998	85.8998
DTU_2016_45	Canada	Ottawa	45.42	-75.69	CAN	North Amer High	4 CAN	79.1554	high	77.20763	37742157	99.5335	67.3625	93.95292	99.58704	82.39025	85.8998	85.8998	85.8998
DTU_2016_46	Switzerland	Liebefeld	46.93	7.42	CHE	Europe & High	4 CHE	80.09504	high	77.20763	8654618	99.838	70.49525	96.54771	99.61471	84.887	89.8016	89.8016	89.8016
DTU_2016_53	China	Guangzhou	23.13	113.26	CHN	East Asia & UpperMidc	3 CHN	72.06164	highMidd	65.7492	1.44E+09	98.5565	81.1	81.68677	97.52467	76.20425	68.6428	68.6428	68.6428
DTU_2016_54	Cote d'Ivoi	Abidjan	5.32	-4.02	NA	NA	NA	CIV	57.56248	low	51.92682	2637825	43.7735	61.65763	73.35314	51.82933	39.16025	55.9624	55.9624
DTU_2016_58	Czech Rep	Brno	49.19	16.61	CZE	Europe & High	4 CZE	81.38791	high	77.20763	10708982	99.795	61.56375	89.77977	98.95133	72.9235	85.5684	85.5684	85.5684
DTU_2016_61	Germany	Berlin	52.52	13.39	DEU	Europe & High	4 DEU	82.4818	high	77.20763	83783945	99.5335	72.44463	93.76762	98.78804	80.70775	84.7016	84.7016	84.7016
DTU_2016_66	Denmark	Lynetten	55.7	12.62	DNK	Europe & High	4 DNK	84.86052	high	77.20763	5792203	99.679	58.85988	94.20269	99.13071	86.78775	90.278	90.278	90.278
DTU_2016_67	Denmark	Avedore	55.63	12.46	DNK	Europe & High	4 DNK	84.86052	high	77.20763	5792203	99.679	58.85988	94.20269	99.13071	86.78775	90.278	90.278	90.278
DTU_2016_68	Denmark	Damhusaa	55.64	12.5	DNK	Europe & High	4 DNK	84.86052	high	77.20763	5792203	99.679	58.85988	94.20269	99.13071	86.78775	90.278	90.278	90.278
DTU_2016_69	Ecuador	Quito	-0.22	-78.51	ECU	Latin Amer UpperMidc	3 ECU	72.53935	highMidd	68.57615	17643060	85.0345	59.19813	77.56379	97.31	77.29775	75.015	75.015	75.015
DTU_2016_70	Ecuador	Galapagos	-0.63	-90.36	ECU	Latin Amer UpperMidc	3 ECU	72.53935	highMidd	68.57615	17643060	85.0345	59.19813	77.56379	97.31	77.29775	75.015	75.015	75.015
DTU_2016_72	Spain	Barcelona	41.38	2.18	ESP	Europe & High	4 ESP	79.46162	high	77.20763	46754783	98.6075	63.8845	94.12762	96.55	85.81375	86.5834	86.5834	86.5834
DTU_2016_73	Spain	Madrid	40.4	29.75	ESP	Europe & High	4 ESP	84.52101	low	84.52101	1.15E+09	92.025	61.53763	84.559	44.38132	53.71345	55.7650	55.7650	55.7650
metadata_species	(2)	metadata_species	(3)																

## Epidemiological/demographic data

## Own data



## Classes of AMR

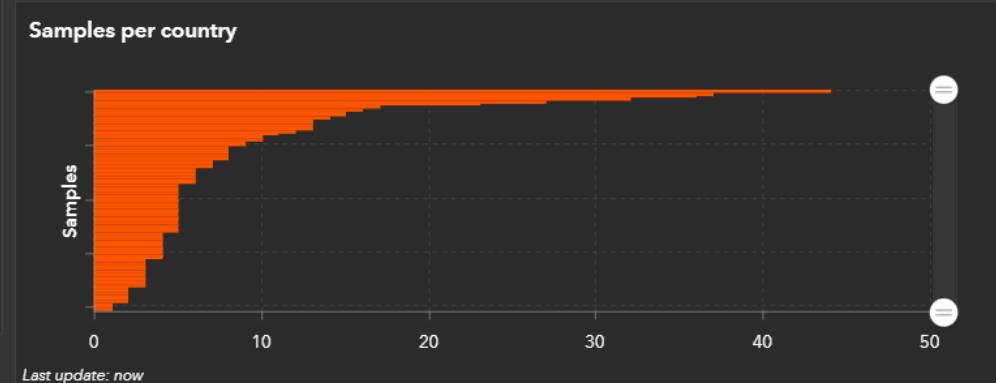
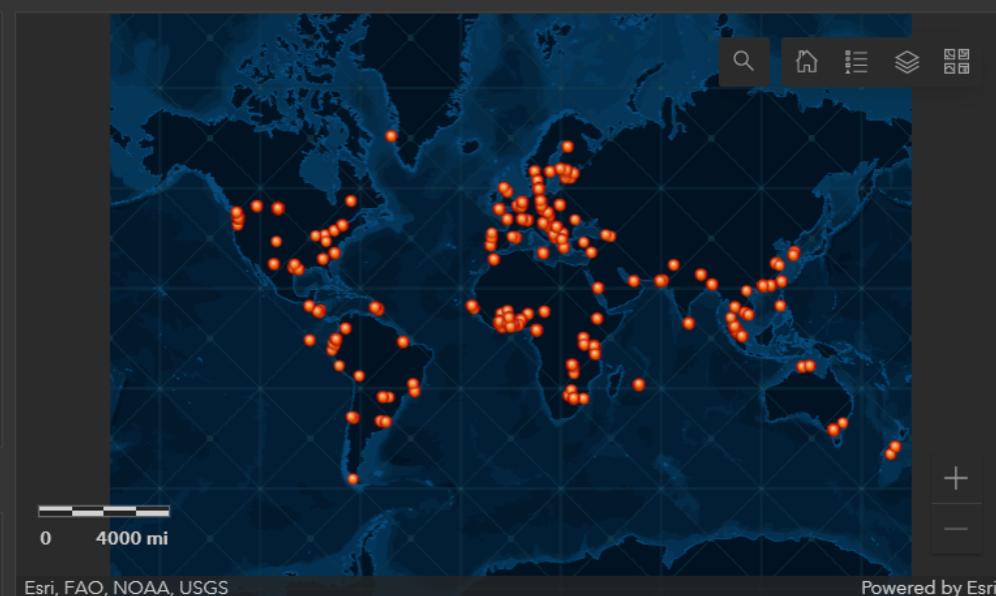
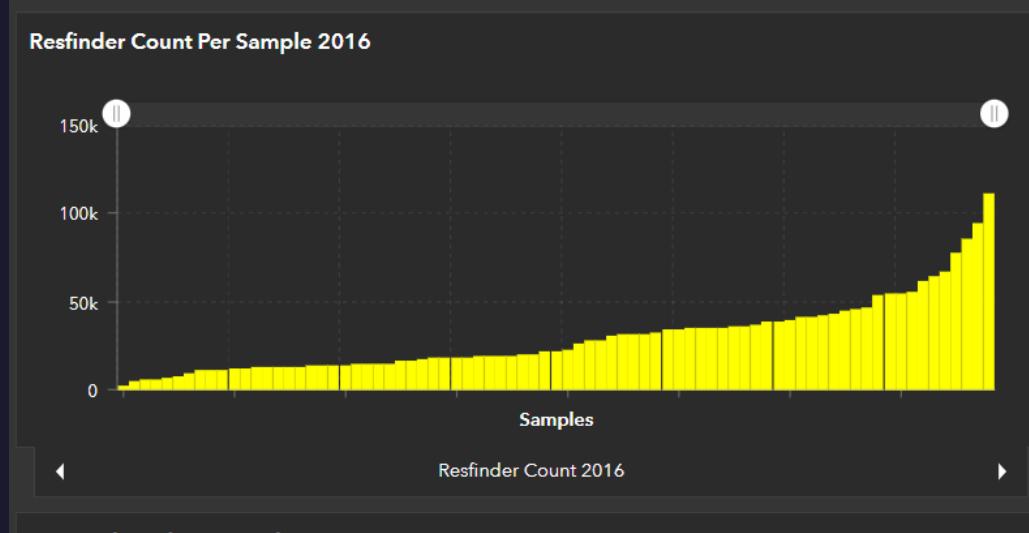
A	B	C	D
country	year	class	fpmk
ALB	16	Aminoglycoside	31955.32
ALB	17	Aminoglycoside	30062.73
ARE	17	Aminoglycoside	67241.45
ARE	18	Aminoglycoside	72546.4
ARG	17	Aminoglycoside	9997.204
ARG	18	Aminoglycoside	17448.68
AUS	16	Aminoglycoside	22699.82
AUS	17	Aminoglycoside	25087.64
AUS	18	Aminoglycoside	18570.99
AUT	16	Aminoglycoside	18128.17
AUT	17	Aminoglycoside	20671.96
AUT	18	Aminoglycoside	20089.6
BEL	17	Aminoglycoside	24318.86
BEL	18	Aminoglycoside	21786.74
BEN	17	Aminoglycoside	76739.81
BEN	18	Aminoglycoside	25558.34
BFA	17	Aminoglycoside	25680.16
BFA	18	Aminoglycoside	25006.62
BGD	17	Aminoglycoside	14786.25
BGD	18	Aminoglycoside	13299.11
BGR	16	Aminoglycoside	32134.56
BGR	17	Aminoglycoside	52197.45
BGR	18	Aminoglycoside	18528.05



## Own data

# Global Monitoring of Antimicrobial Resistance using Urban Sewage Samples (Phase II): Global Sewage Surveillance Project (...

Data measures: total counts silva bacterial load, resfinder, and metadata



### About:

This dashboard was developed by Nermin Ghith, Postdoc Researcher at the Research Group for Genomic Epidemiology, Division for Global Surveillance, National Food Institute, Technical University of Denmark (DTU Food).

**Go to project StoryMaps:**  
<https://storymaps.arcgis.com/stories>

### Country/City/Municipality/Sample ID

- Albania  
• Tirana  
2016\_pil\_alb\_01
- Albania  
• Tirana  
2017\_jun\_alb\_01
- Albania  
• Tirana  
2017\_nov\_alb\_01
- Argentina  
• Buenos Aires  
2017\_nov\_arg\_01
- Argentina  
• Buenos Aires  
2018\_jun\_arg\_01
- Argentina  
• Buenos Aires  
2018\_nov\_arg\_01

Last update: now

# Opportunities and challenges

- Large collection of samples (and countries)
- Public interest and international network of collaborators
- WGS technology but metagenomics data (bacterial taxa)
- Too many zeroes (up to 73-80% at the species and gene levels)
- Aggregate resistance genes into phenotypes (antibiotics) and phenotypes into classes of AMR (zeroes are around 6%)
- Selection and interpretation of data analyses



# Methods

- The Global Sewage Surveillance project consortium collects untreated sewage samples to assess the global burden of AMR.
- Site-specific samples (560 and 774) were collected between 2016 and 2018 and came from 233 cities (102 countries).
- Samples were sequenced to obtain information on 19 classes of antimicrobial resistance.
- The read counts per class of AMR were transformed using the centered log-ratio (clr) to obtain the outcome measures.



# Methods

- Applying random effect models, the contextual effect of the city was quantified using the variance partition confident (VPC) by adding the random effects of the city in the analyses.
- The model variance was estimated as the median of the posterior distribution obtained by the Markov Chain Monte Carlo (MCMC) method.
- SDG indices (determinants) were added into the models to evaluate the study aim.
- Data were extracted from the World Bank and the Global Burden of Disease data portals.
- Commercial data bought from GIDEON and IQVIA, among other resources.
- Use proportional change in variance (PCV) to assess the effect of the determinants (each in a separate model)



# Methods

## Determinants in the models

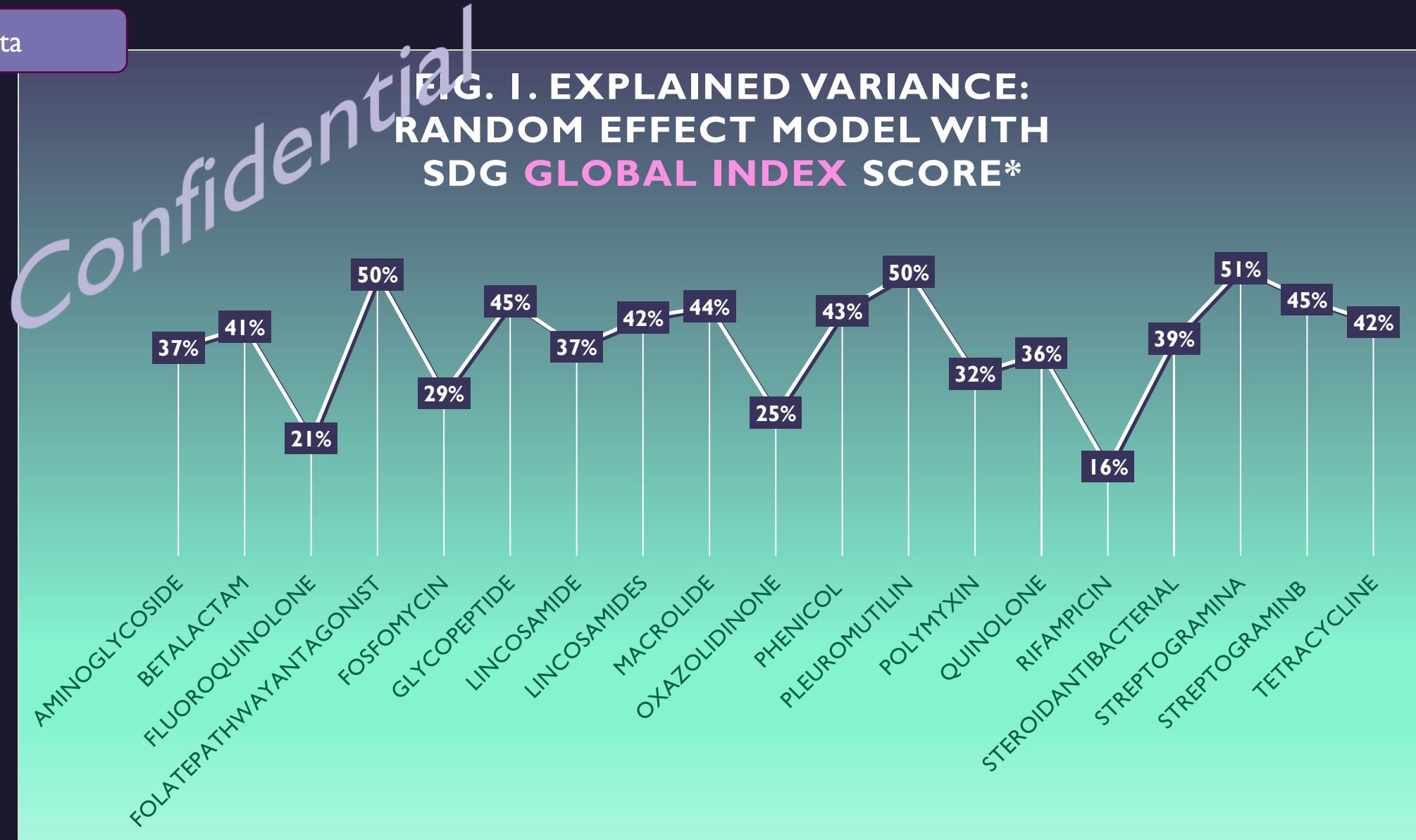
- SDG global index: covers all the 17 goals
- Universal Health Coverage (UHC)
- Vaccine Cover
- WaSH Mortality: \* Mortality rate attributed to unsafe water, unsafe sanitation and lack of hygiene (exposure to unsafe Water, Sanitation and Hygiene for All (WASH))
- Access to clean water
- GBD-SDG health: index composed of 41 health indicators



# Findings

- Adding the SDG global index into the random intercept models explained around 6%-55% of the between city variation in AMR .
- Similar findings obtained by adding the indices on Universal Health Coverage (UHC), Vaccine Cover,WaSH Mortality and access to clean water indicators into the models.
- In all the models, the largest reduction in between city variation in the burden of AMR occurred for folate pathway antagonist, Phenicol, and Tetracycline classes.
- The least reduction was for polymyxin and quinolone classes.

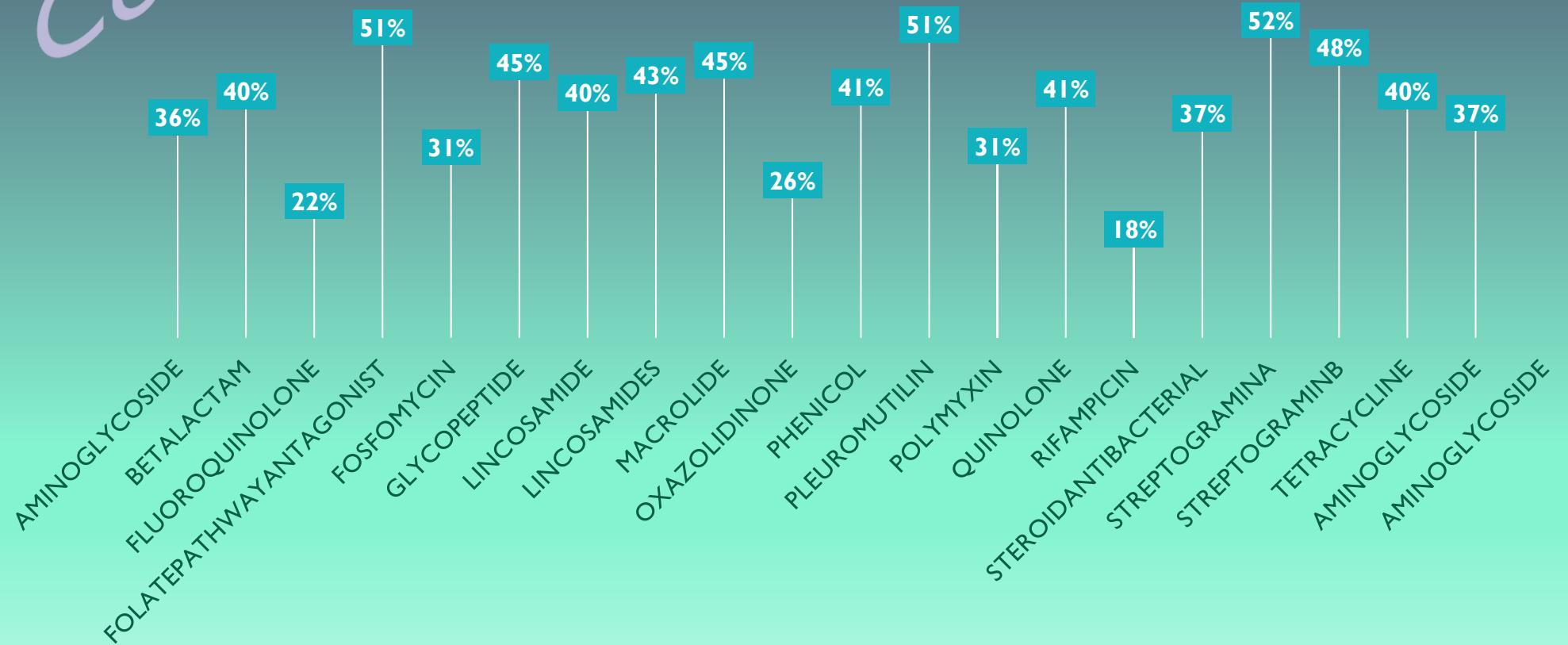




\*The index covers all the SDGs 17 goals



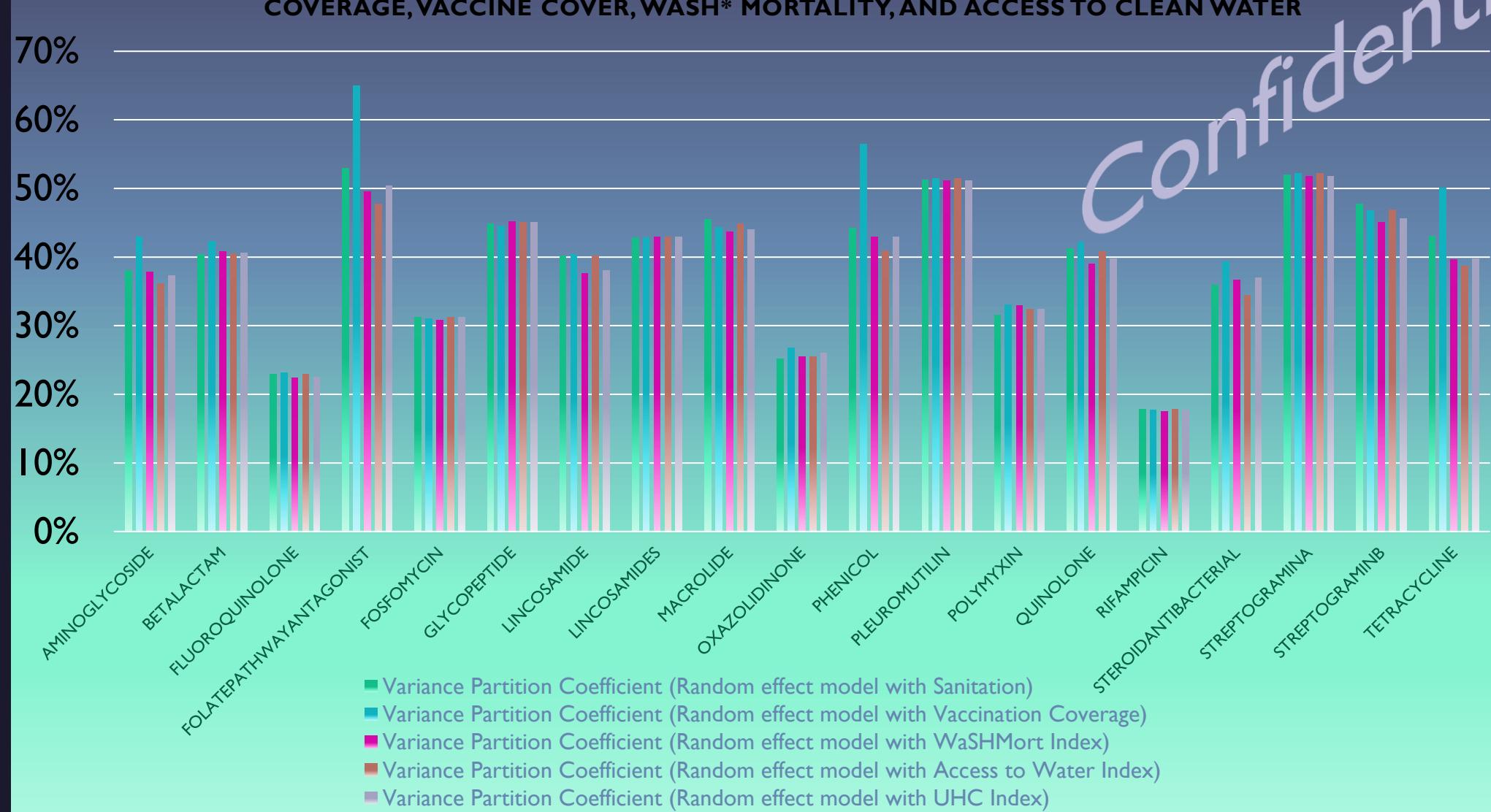
Confidential

**FIG. 2. EXPLAINED VARIANCE:  
RANDOM EFFECT MODEL WITH  
SDG HEALTH INDEX\***

\* The index covers 41 health indicators (health index)



**FIG. 3. EXPLAINED VARIANCE:  
RANDOM EFFECT MODELS WITH INDICES ON SANITATION, UNIVERSAL HEALTH  
COVERAGE, VACCINE COVER, WASH\* MORTALITY, AND ACCESS TO CLEAN WATER**



\* Mortality rate attributed to unsafe water, unsafe sanitation and lack of hygiene (exposure to unsafe Water, Sanitation and Hygiene for All (WASH))





# Sustainability Development Goals and biomarkers of Antimicrobial Resistance: differential abundance analyses

Nermin Ghith, Hannah-Marie Martiny, Patrick Munk, Amalia Bogri, and Frank Møller Aarestrup

Research Group of Genomic Epidemiology, Technical University of Denmark (DTU)



Research group of Genomic Epidemiology  
Technical University of Denmark  
Mobile: +4528190249  
[research@dtu.dk](mailto:research@dtu.dk)  
<https://www.ncbi.nlm.nih.gov/nuccore/50059311/>

Presentation date: 6/12/2022  
PRESENTATION (Publication) NUMBER 2581

## Revised abstract

### Background:

Inequality in global health outcomes such as antimicrobial resistance (AMR) is shaped by macro determinants of health within and between countries.

Therefore, this study aims to assess the potential relationship between the country's development level - progress towards achieving the sustainability development goals (SDG) and the burden of AMR.

The study identifies resistance genes, phenotypes (antibiotics), classes of antimicrobial resistance and bacterial taxa that are differentially abundant between the sampling sites within countries, and compliance with SDGs goals that might explain variation in the global burden of AMR between countries.



### Methods:

The Global Sewage Surveillance project consortium collects untreated sewage samples to assess the global burden of AMR. Samples were sequenced to obtain data on the microbial compositions within site-country covering information on resistance genes and bacterial taxa. Samples were collected between 2016 and 2018 and came from 233 cities (102 countries).

A combination of models is developed to describe and characterize the data on AMR in association with SDG indices and goals. A number of SDG indices were added to the analyses to evaluate the study aim. Data on SDGs indices were extracted from the World Bank and the Global Burden of Disease data portals.

### Results:

The microbial communities across sites differ from one another, which are attributed to site-country development levels (e.g., measured with SDG index and goals). The study identified a number of differentially abundant taxa at different levels of the bacterial taxonomy as well as differentially abundant genes, phenotypes and classes of AMR.

### Conclusions:

Further interventions focusing on improving adherence to SDG goals have the potential for elevating inequality in the global burden of AMR. Yet, a complementary assessment for additional sampling locations is required to get data that are more granular on the global and local burden of AMR and to determine the most appropriate course of action.

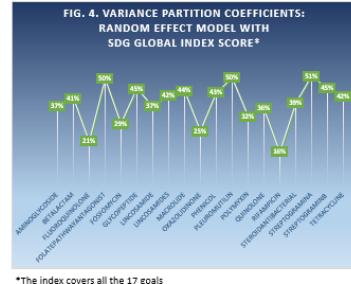
## Methods

Metagenomics samples were sequenced to obtain information on 19 classes of antimicrobial resistance, 65 phenotypes, 562 resistance genes and bacterial taxa.

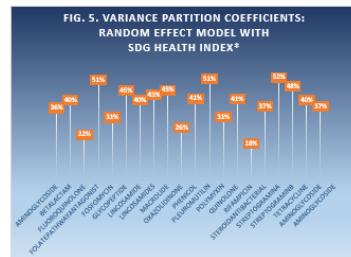
To assess the study aim, the findings were triangulated using the Wilcoxon-Mann-Whitney test (WMW), negative binomial models (NB), random effect models, compositional data analysis, and analysis of compositions of microbiomes with bias correction (ANCOM-BC or ABC).

Applying random effect models, the contextual effect of the city was quantified using the variance partition coefficient (VPC) by adding the random effects of the city in the analyses.

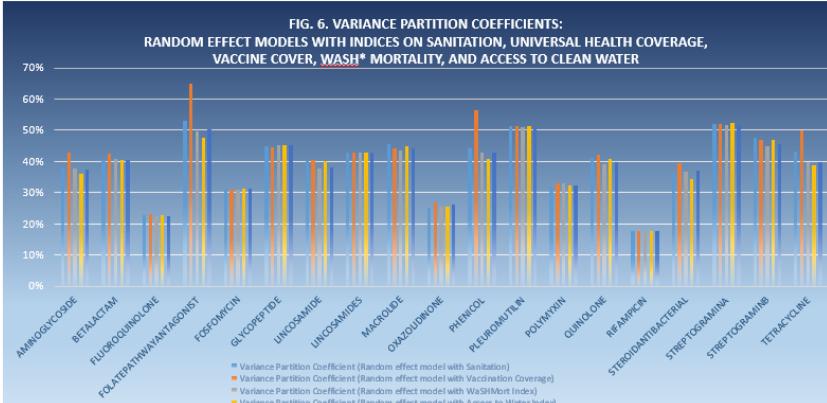
A number of SDG indices were added into the models to evaluate the study aim. Data on SDGs indices were extracted from the World Bank and the Global Burden of Disease data portals. The model variance was estimated as the median of the posterior distribution obtained by the Markov Chain Monte Carlo (MCMC) method.



\*The index covers all the 17 goals

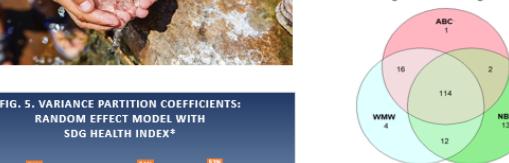


\* The index covers 41 health indicators



\* Mortality rate attributed to unsafe water, unsafe sanitation and lack of hygiene (exposure to unsafe Water, Sanitation and Hygiene for All (WASH))

Fig. 1 Resistance genes



Results

The microbial communities across sites differ from one another, which is attributed to site-country development levels (e.g., measured with SDG index and goals). The study identified 114 resistance genes, 35 phenotypes and 11 classes (Figs. 1-3) that are differentially abundant between countries.

Classes of AMR that have the smallest p-values across the random effect model analyses are: Aminoglycoside, Folate pathway antagonist, Lincosamide, Macrolide, Oxazolidinone, Phenicol, Pleuromutilin, Polymyxin, Streptogramin-A, Streptogramin-B, and Tetracycline.

Adding the SDG global index and SDG health scores into the random intercept models explained around 18%-52% of the between city variation in total AMR (Figs 4 and 5). Similar findings obtained by adding the indices on Sanitation, Universal Health Coverage (UHC), Vaccine Cover, WaSH Mortality, and Access to water individually into the models. The variance partition coefficients (VPC) values are between 18%-65% (Fig 6).

In random effect models (Figs 4-6), considering the SDG indices, the largest reduction in between city variation in the burden of AMR occurred for folate pathway antagonist, pleuromutilin and streptogramin-A classes, and the least reduction (still quantitatively large) was for fluoroquinolone, oxazolidinone and rifampicin classes.

## Conclusions

Actions focusing on sustained compliance with the SDG goals have the potential for elevating inequality in the global burden of AMR. There is a need for additional data that cover more sampling locations to develop more comprehensive global monitoring of AMR, and to determine the most appropriate course of action.



# Conclusion

- Further **interventions** focusing on improving adherence to SDG goals have the potential for elevating the global burden of AMR.
- Yet, a complementary assessment for **additional locations** is required to get more granular data on the global burden of AMR and to determine the most appropriate course of action.
- Consider other **analytical frameworks**: e.g., log linear transformation of FPKM and offset for the library size



# Teaching: New trends in surveillance



Surveillance and Epidemiology of infectious Diseases

MSc in Bioinformatics – master track in Infectious Disease Informatics

Sewage II 2024 - RStudio

Edit Code View Plots Session Build Debug Profile Tools Help

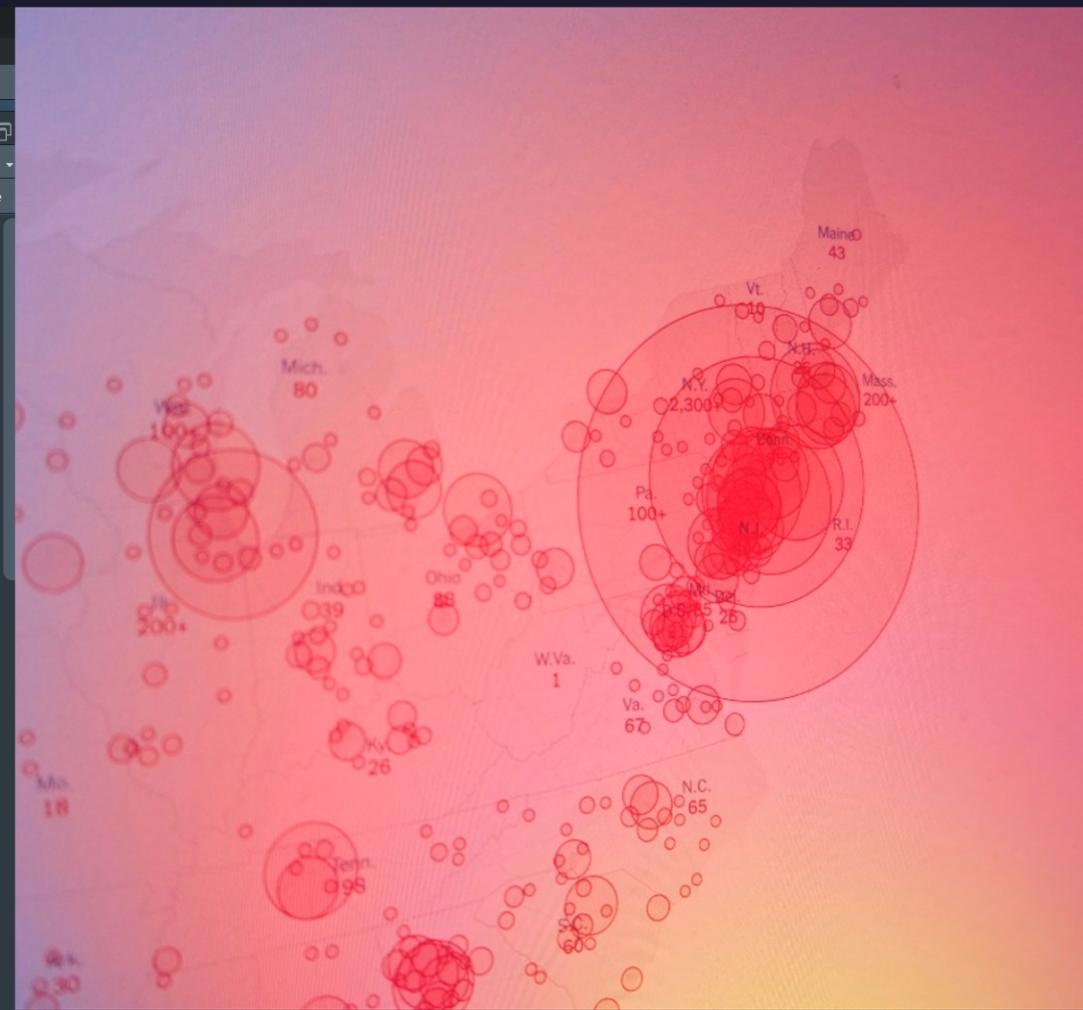
Go to file/function Addins

Data processing bacteria taxa.Rmd SDGsVIZ.qmd SDGs2024\_Q2.2\_AllData\_After\_20240729 SDG.Rmd

Knit on Save ABC Knit Run Outline

Source Visual

```
1 ---  
2 title: "Data processing class Pheno u90 total"  
3 author: "Nermin Ghith"  
4 date: "2024-04-17"  
5 output: html_document  
6 ---  
7  
8 # view the data  
9  
10 ```{r setup, include=FALSE}  
11 knitr::opts_chunk$set(echo = TRUE)  
12  
13 # # List of file names  
14 # file_names <- c("bacteriaLoadPerSample", "resfinder.counts.phenotype", "resfinder.counts.uc90",  
15 # "resfinder.fpkm.class", "resfinder.fpkm.phenotype",  
16 # "resfinder.fpkm.uc90", "total.counts.silva+resfinder")  
17 #  
18 # # Loop through each file name  
19 # for (file in file_names) {  
20 #   # Construct the file path  
21 #   file_path <- paste0("C:/Users/nermi/Dropbox/DTU/Global determinants of AMR/Sewage II  
2024/Data/kma_quantification/products/", file, ".txt")  
22 #  
23 #   # Read the file and assign it to a unique variable in the global environment  
24 #   assign(file, read.delim(file_path))  
25 # }  
26 #  
27 # # manually view each data frame in RStudio with view(resfinder.counts.phenotype), etc.  
28  
29  
30  
31  
32 # Set the path to the data directory  
33 directory_path <- "C:/Users/nermi/Dropbox/DTU/Global determinants of AMR/Sewage II"
```

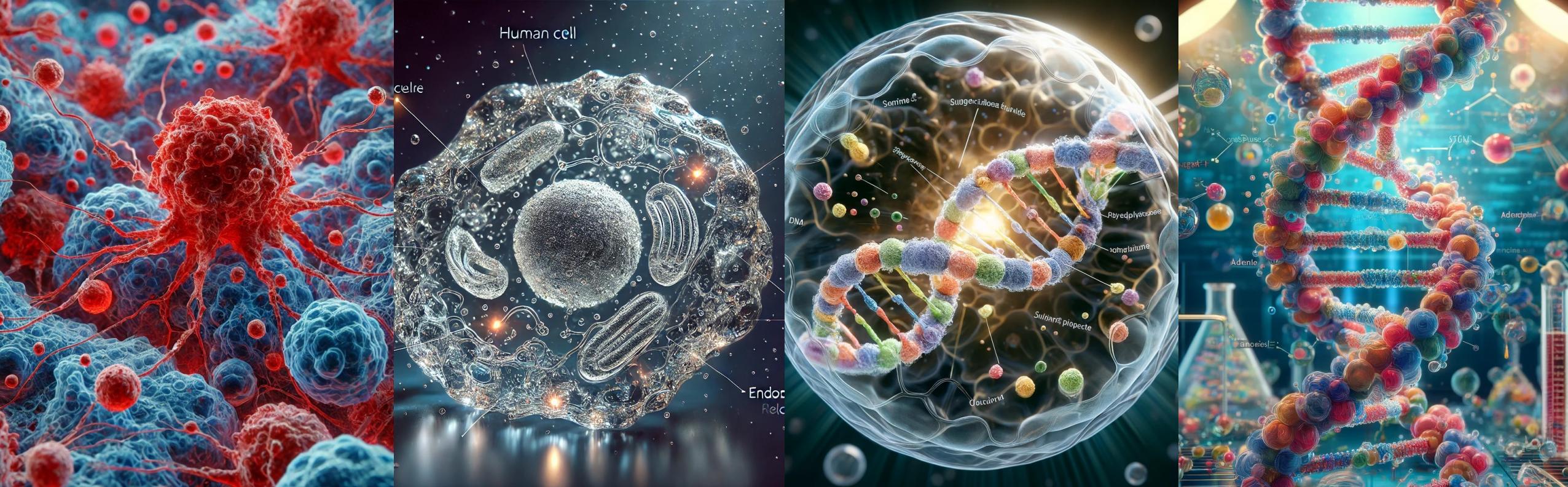


Omics-Based Clinical and Population Studies (session 1-2)

## AI-Driven Patient Care

Transforming the Experience





# Cancer Epi- omics cohort

Epi-omics And Drug Use: Identification Of Late Effects  
Trajectories  
  
Variants of Uncertain Significance in Cancer and Patient Risk  
Profiles



## SALICCS PROGRAMME EXPERIENCE: SCIENTIFIC COORDINATOR FOR THE PROGRAM

- My experience - the Nordic cross-border research programme SALiCCS, a collaboration with the Danish Cancer Society, the Finnish Cancer Registry, and Karolinska Institutet. This programme investigates late treatment effects and psychosocial conditions in childhood cancer survivors across Denmark, Finland, and Sweden.
  - **Cohort Design:** SALiCCS is a registry-based matched cohort design including five-year survivors diagnosed between 1971-2011, along with age-, sex-, and country-matched comparisons and siblings.
  - **Outcomes of Interest:** The programme examines morbidity and disease progression, mental disorders, education, employment, family life, and other related outcomes.
  - **Data Linkage:** Data from national patient, cancer, and medication registries were used for individual-level data linkage.
  - **Core Population:** The core population includes 21,292 five-year survivors, 103,303 population comparisons, and 29,644 siblings. Common diagnoses include central nervous system tumors, leukemias, and lymphomas.



# Research domains: SALICCS: Nordic research projects



## Late Effects

Create hiigh-resselovisions animaed illustsation



## Treatments: Use of medications (whole profile)

- Compositions of treatments (chemo, radio, immunotherapies, etc.)
- Characterization of compositions (e.g., morbidity, stage of disease, demographics, etc.)

## Quality of life: national survey

- Changes in life course: compositions of life events, activities, etc
- Social and health outcomes in survivors: compositions of patient socioeconomic and family context, trajectories of health late effects

## COVID-19

- in children, young adults, and survivors of childhood cancer

## Linking epi data to omics

- E.g., Children with leukemia (samples from healthy siblings' vs sick children)



# Danish Data for my projects: DST, RKKP and SDS

Date Created: January 20, 2025  
Last Modified: January 20, 2025

Purpose:  
Link multiple Danish health registries including LPR2/LPR3 (patient registry),  
medication registry, vaccination registry, and population registry.

Input Datasets:

1. LPR.KONTAKTER - Hospital contacts from LPR3
2. LPR.DIAGNOSEN - Diagnoses from LPR3
3. LPR.DIAG\_INDL - Historical diagnoses from LPR2
4. MED.LMDB202312 - Medication data (December 2023 extract)
5. MED.LMDB202306 - Medication data (June 2023 extract)
6. VAC.EPIDDV\_VACCINEDATA - Vaccination data
7. BEF.BEF202312 - Population registry (December 2023 extract)

Output Datasets:

1. WORK.FINAL\_COHORT - Final linked dataset
2. WORK.PATIENT\_TIMELINE - Detailed patient event timeline

Notes:

- LPR2 data cutoff: 30NOV2013
- LPR3 contains data from 2018/2019 onwards
- Population database contains only living citizens
- Multiple records per patient exist in all datasets except population

## Indholdsfortegnelse

- 1 [Intro](#)
- 2 [Data flow](#)
  - 2.1 [Data kilder](#)
  - 2.2 [Data ansøgning](#)
- 3 [Hvordan læser jeg dokumentet?](#)
- 4 [Befolningen](#)
  - 4.1 [Det centrale personregister \(CPR-registeret\)](#)
- 5 [Ind-/udvandringer](#)
- 6 [Død og dødsårsager](#)
- 7 [Cancerregisteret](#)
- 8 [Uddannelse](#)
- 9 [Indkomst](#)
- 10 [Arbejdsmarkedstilknytning](#)
- 11 [Landspatientregisteret](#)
  - 11.1 [LPR2](#)
    - 11.1.1 [Indlæggelser](#)
    - 11.1.2 [Privathospitaler](#)
    - 11.1.3 [Psykiatriske indlæggelser](#)
    - 11.1.4 [Fødsler fra LPR](#)
    - 11.1.5 [Medicinsk fødselsregister](#)
    - 11.1.6 [Fertilitetsdatabasen](#)
  - 11.2 [LPR3\\_F](#)
- 12 [Register over In Vitro Fertilitetsbehandling \(IVF\)](#)
- 13 [Landsdækkende Register for Patologi](#)
- 14 [Det Psykiatriske Centrale Forskningsregister](#)
- 15 [Sygesikringsregisteret](#)
- 16 [Register for Udvalgte Kroniske Sygdomme og svære psykiske lidelser](#)
- 17 [Lægemiddeldata](#)
  - 17.1 [Lægemiddeldatabasen](#)
  - 17.2 [Sygehusmedicinregisteret](#)
- 18 [Laboratoriedata](#)

@NerminGhith

## Variabelliste for DST

KB projektdatabase 704075

5/26/24, 11:56 PM

706024\_project proposal\_20240526\_v2

# APPENDIX 1 – Appendix documents available in application

Project proposal 706024 (2):  
Adult Life after Childhood  
Cancer in Scandinavia(2018-  
DCRC-0044)

FILE NAME	ALIAS	ATTACHED TO	DATE
Additional data from Denmark_ng.pdf		Data fra projektdatabasen (704075) ( POP B )	06.12.2023
Additional data from Sweden.pdf		Nordic data ( POP B )	03.11.2023
Additional variables from Finland.pdf		Nordic data ( POP B )	03.11.2023
EpiDDV_Vaccinedata.xlsx		Data from SDS ( POP B )	08.12.2023
New 2023-05-01_Variabeloversigt_DBCR.XLSX		Data from RKKP ( POP B )	05.12.2023
New FOMY Parental custody.docx		Data fra projektdatabasen (704075) ( POP B )	23.09.2023
New RKKP Variabeloversigt_DACOVID.xlsx		Data from RKKP ( POP B )	27.10.2023
New RKKP Variabeloversigt_DID.xlsx		Data from RKKP ( POP B )	27.10.2023
New SDS covid-19 testdata pr. case.xlsx		Data from SDS ( POP B )	07.12.2023
New__LPR.xlsx		Data fra projektdatabasen (704075) ( POP B )	06.12.2023
New_LMDB.xlsx		Data from SDS ( POP B )	06.12.2023
Parental information.xlsx		Data fra projektdatabasen (704075) ( POP B )	05.12.2023
Variables on cohabitation, leaving home, marital status.xlsx		Data fra projektdatabasen (704075) ( POP B )	05.12.2023
Variables on education.xlsx		Data fra projektdatabasen (704075) ( POP B )	05.12.2023
Variables on occupation, work, income.xlsx		Data fra projektdatabasen (704075) ( POP B )	05.12.2023
Variables on parenthood.xlsx		Data fra projektdatabasen (704075) ( POP B )	05.12.2023
Variables on social security benefits.xlsx		Data fra projektdatabasen (704075) ( POP B )	05.12.2023

# Drug use

Use of prescription drugs among childhood cancer survivors – a population-based study of late effects in Nordic countries

## MAPPING GENERAL USE

The overall use of prescription drugs and prevalence of polypharmacy in childhood cancer survivors including all fourteen main groups of prescription drugs and subgroups (Study 1)



## ANTIBIOTICS

The use of antibiotics in childhood cancer survivors compared to that in the background population  
(Study 2)



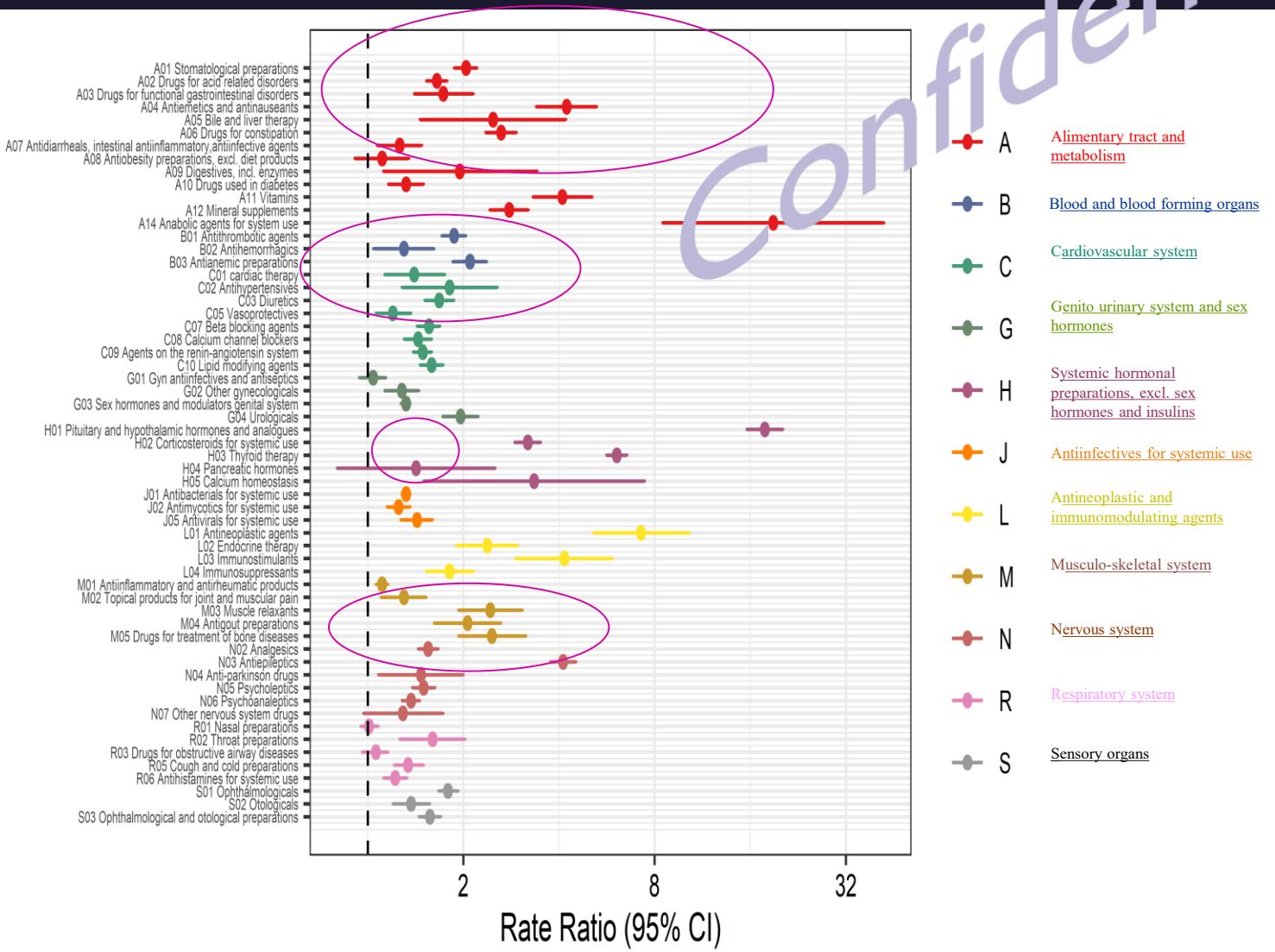
## NERVOUS SYS. DRUGS

The use of analgesics , antidepressants, and antipsychotics prescribed and especially potent analgesics or opioids in childhood cancer survivors compared to that in the background population (Study 3)

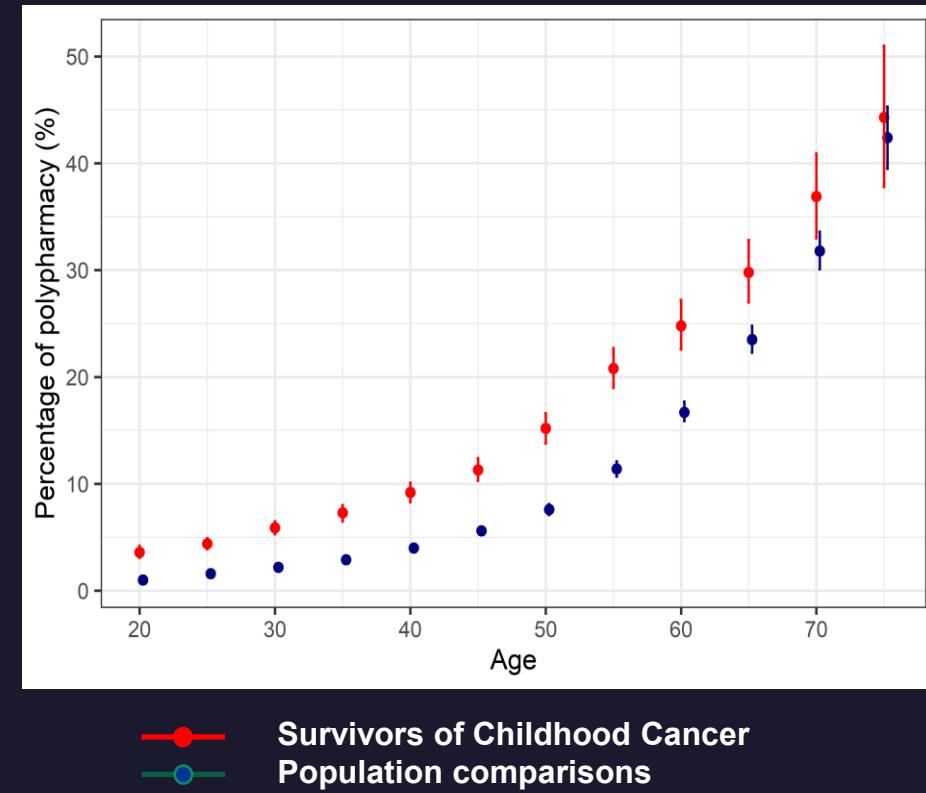


## Own data

**Figure 1. Rate ratios and 95% CIs for prescriptions within the 65 ATC 2<sup>nd</sup> level groups of prescription drugs (dispensed more than 100 times among survivors), comparing survivors of childhood cancer with population comparisons**



**Figure 2. Prevalence and 95% confidence intervals in childhood cancer survivors and population comparisons with polypharmacy**



# Proposed new cohort: Linking Epi-omics and Drug Data For The Identification Of Late Effects Trajectories

## The Recent Research Proposals: Gene Variants In Childhood Cancer

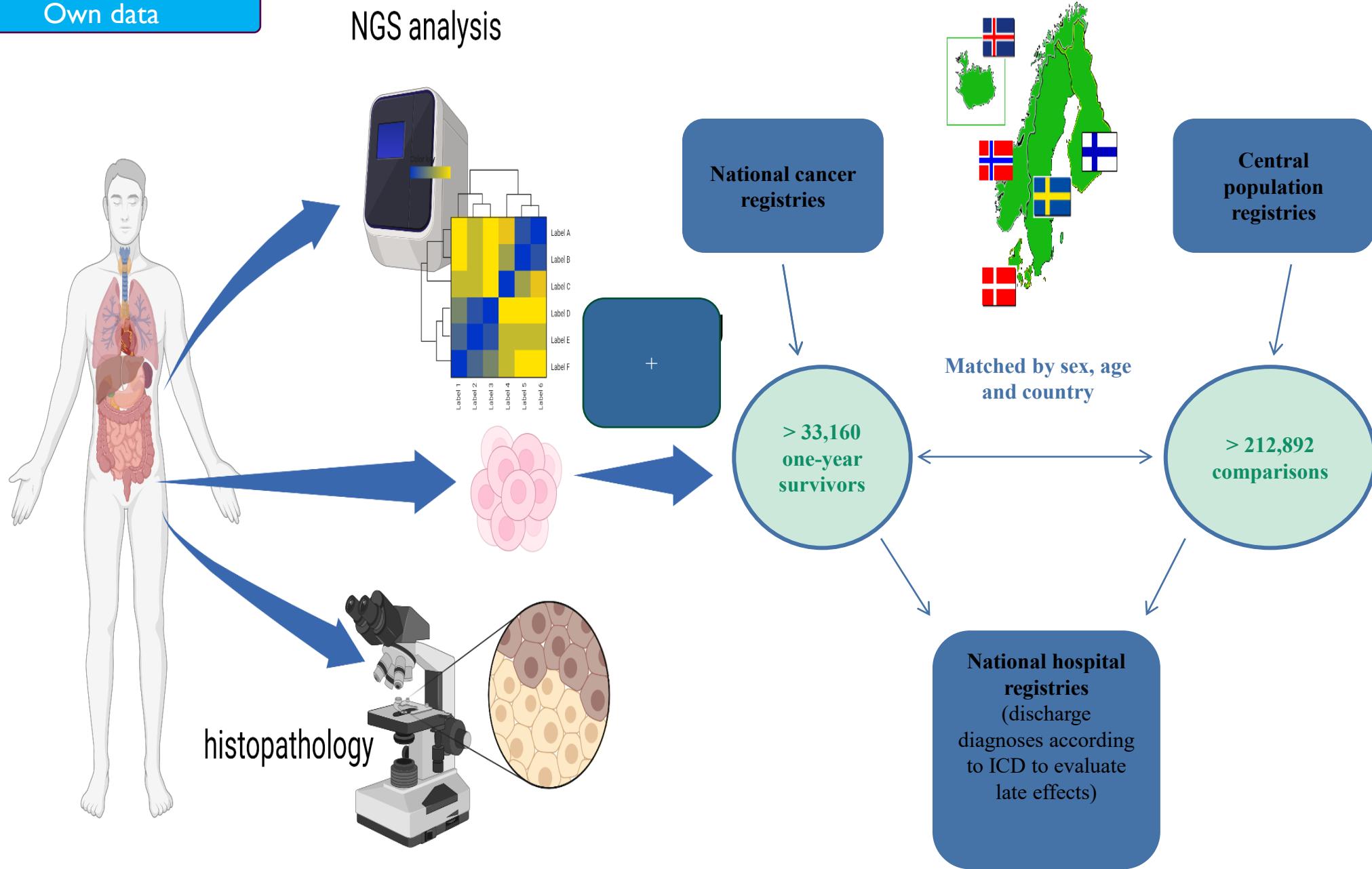
### Key Techniques and Approaches

- **Large-Scale Record Linkage:** Utilizing national health outcome registers in nationwide population-based settings to link various large-scale records.
- **Characterized Variants and Clinical Data:** Leveraging data from registries to enhance understanding of tumor evolution, triggers, and risk factors for second cancers and late effects and improve accuracy in treatment response predictions and survival estimates.
- **Comprehensive Dataset Development:** Using national and regional cancer registries to develop a comprehensive dataset on children, teenagers, and young adults with cancer.
- **Population-Level Analysis:** Conducting large-scale analyses to explore the relationship between genetic variants and the onset of cancer, survival, treatment resistance, and late effects.
- **Follow-Up Health Data:** Utilizing medical records, prescription and medication details, and follow-up health data from national registries for survivor comparisons.



## Own data

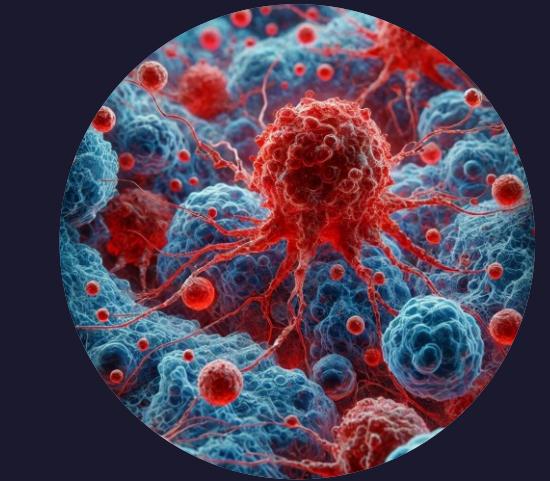
# NGS analysis



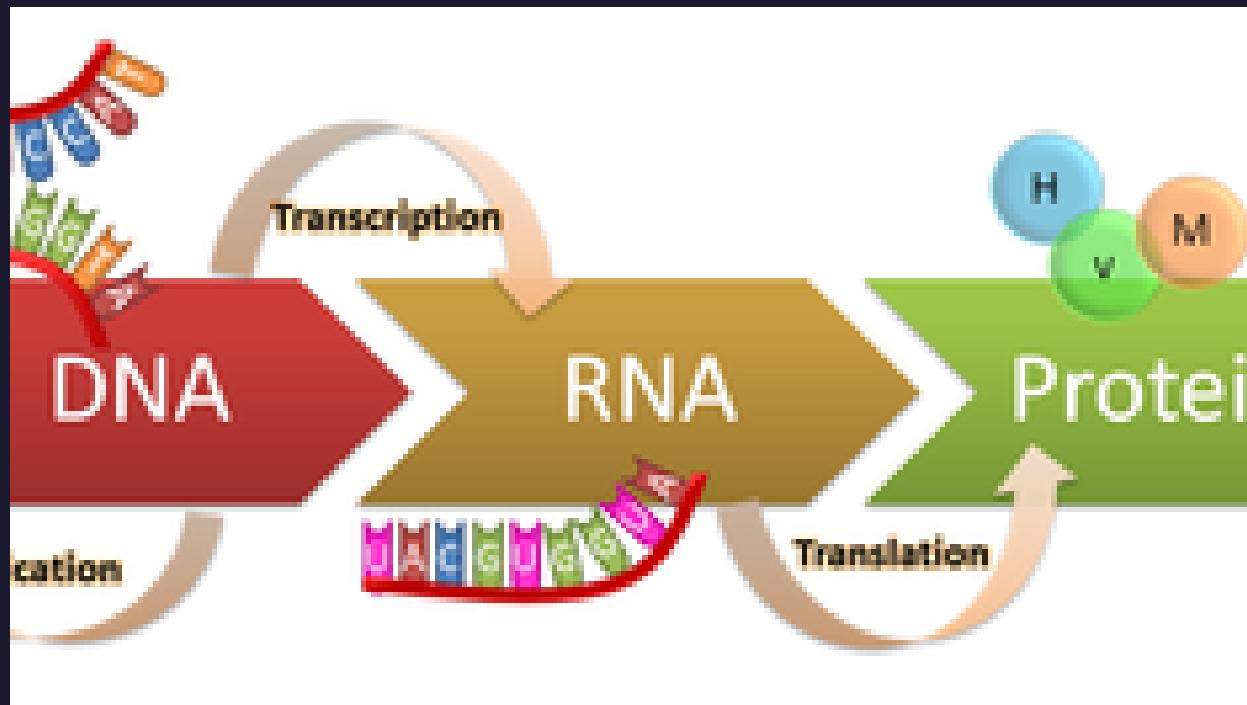
# New Epi. Cohort to Investigate: Variants Of Uncertain Significance In Rare Types Of Cancer

Characterization and **stratification** of patient morbidity clusters, risk of secondary cancer, prognosis, and survival

- Link **transcriptomic** data with clinical and population data in the national registries.
- Compare patients with their **siblings** and the **general population**



# Transcriptomics



- “**Transcriptomics** is the study of the structure, function, and evolution of the **transcriptome** (i.e., the entirety of **RNA transcripts** produced by the genome) of a given organism or community of organisms under a variety of conditions”.
- “An **RNA transcript** is the **RNA strand** that is produced when a gene is transcribed.”

Sources: <https://toolkit.ncats.nih.gov/glossary/rna-transcript#:~:text=An%20RNA%20transcript%20is%20the,is%20translated%20into%20a%20protein>.

<https://www.mrgscience.com/topic-27-dna-replication-transcription-and-translation.html>

<https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/transcriptomics>

VUS definition:

"A change in a gene's DNA sequence that has an unknown effect on a person's health. There is usually not enough information about a variant of uncertain significance to know whether it increases a person's risk of developing a disease, such as cancer. Also called unclassified variant, variant of unknown significance, and VUS."

Source:

<https://www.cancer.gov/publications/dictionaries/cancer-terms/def/variant-of-uncertain-significance>

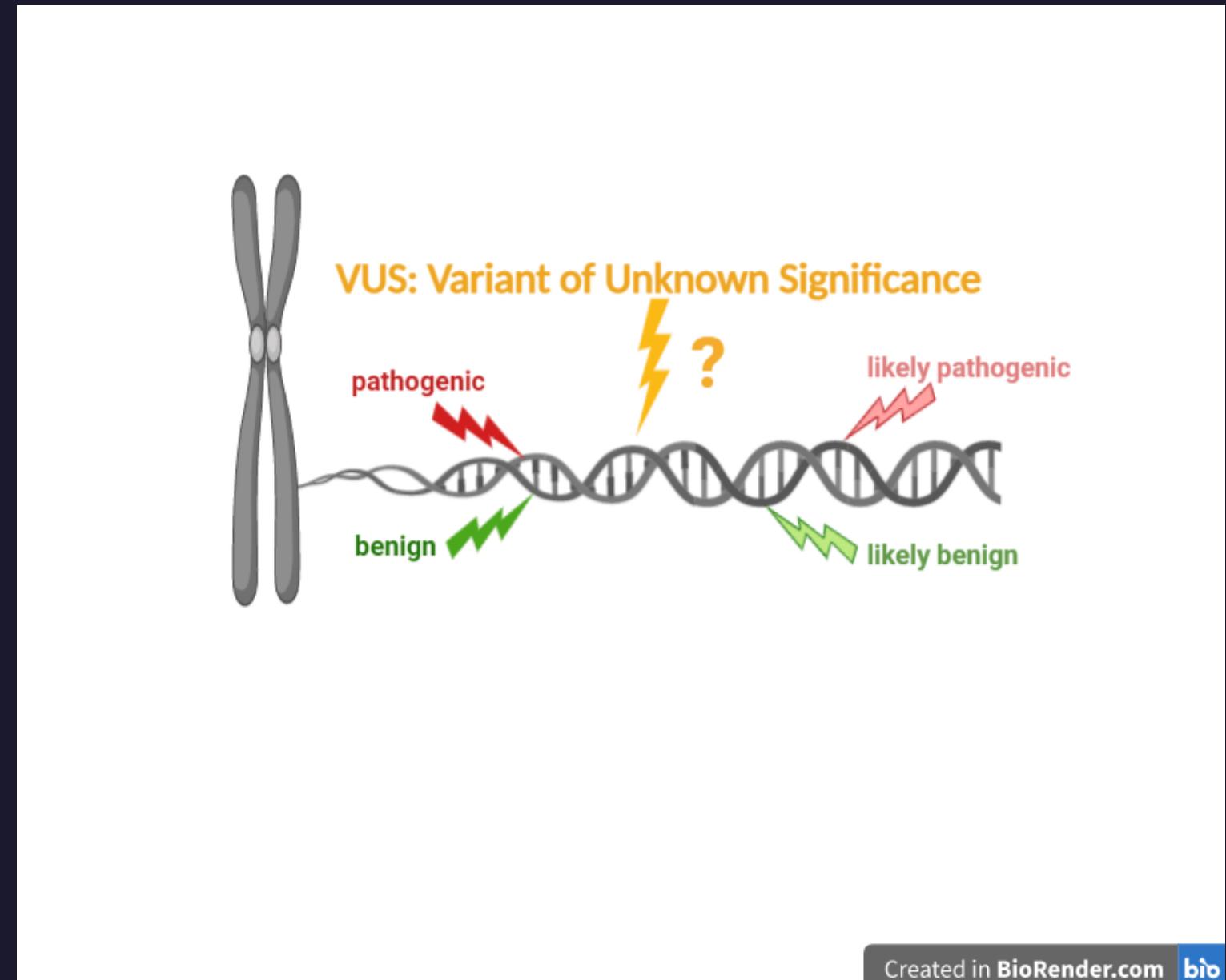
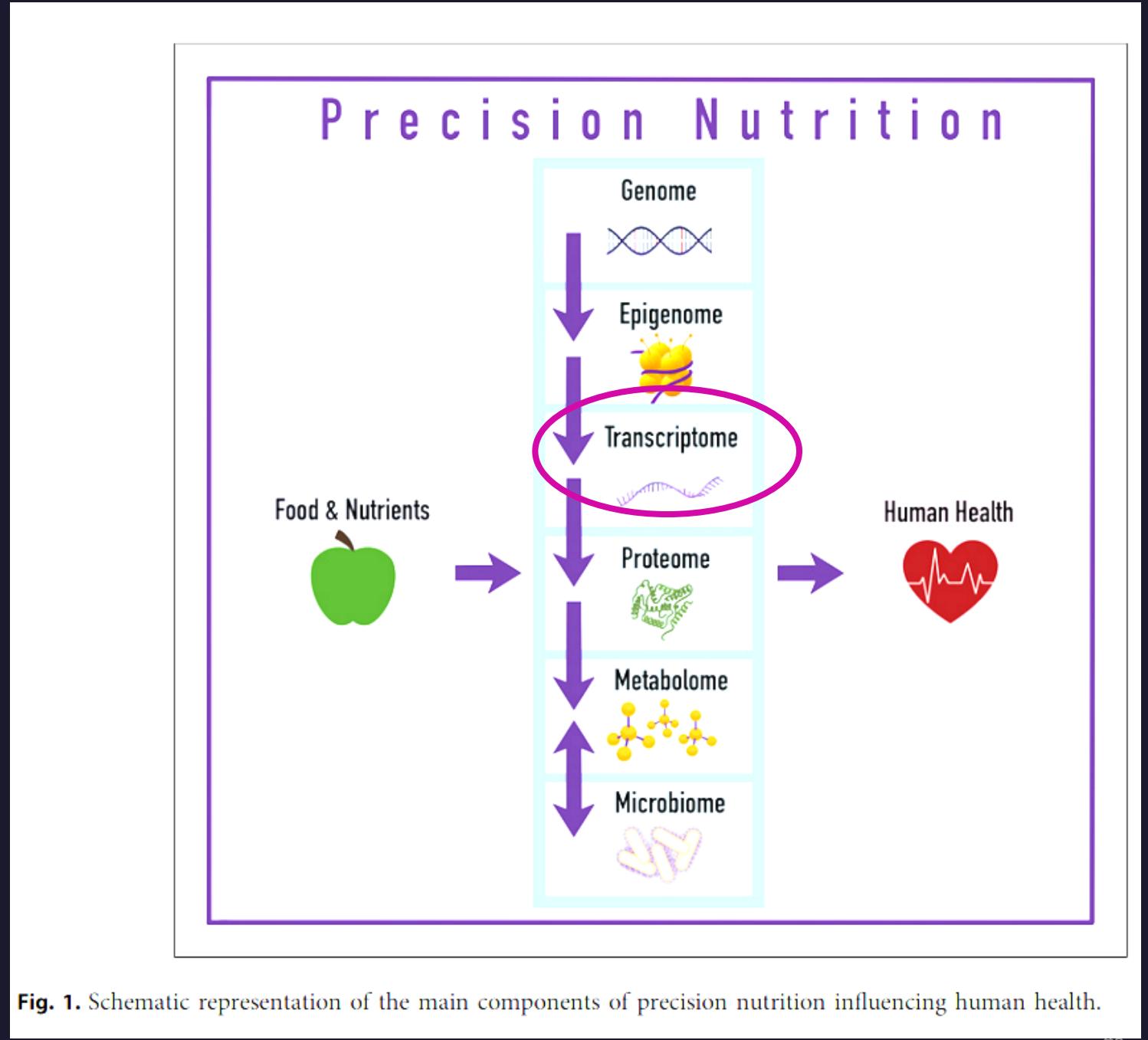


Figure source: <https://curesyngap1.org/blog/does-your-genetic-report-contain-a-variant-of-unknown-significance-vus-in-syngap1/>

Created in BioRender.com

Source:  
**Precision Nutrition for Cardiovascular  
Disease Prevention**

- January 2023
- Lifestyle Genomics 16(1)
- DOI:
- [10.1159/000529054](https://doi.org/10.1159/000529054)
- License
- CC BY-NC 4.0



# Research domains: SALICCS Nordic research projects


AARHUS  
UNIVERSITY

**Use of prescription drugs in Danish and Swedish childhood cancer survivors: A population-based cohort study**

Nermin Ghith and Jeannette F. Winther, MD, DMSc\*

\* On behalf of SALICCS research program collaborators

**Aim**

The aim of this cohort study was to provide a detailed yet comprehensive overview of the use of prescription drugs in childhood cancer survivors diagnosed in Denmark and Sweden as well as to investigate the prevalence of polypharmacy.

**Methods**

Based on unique data from the Nordic population-based research program SALICCS, all children with cancer before the age of 20 years in Denmark (1945-2000) and Sweden (1956-2011) were identified. The study population consisting of 16,742 five-years childhood cancer survivors identified in the nationwide cancer registries and classified according to the International Classification of Childhood Cancer, with 77,701 population comparisons (without cancer) randomly selected from the population registers and 23,105 sibling comparisons. Information on drug prescriptions was obtained by linking the three cohorts to the nationwide prescription registers in Denmark and Sweden with information on all drugs dispensed at pharmacies using the hierarchical Anatomical Therapeutic Chemical (ATC) classification system. For each dispensing, information was obtained on the ATC-code at 5<sup>th</sup> level (chemical/therapeutic groups of prescription drugs). Rates ratios (RRs) and 95% confidence intervals (CIs) for filled drug prescriptions were calculated in marginal rates models allowing each individual to have several prescriptions during follow-up. Relative rates of prescriptions were calculated for the 11 ATC 1<sup>st</sup> main level and each of the 65 ATC 2<sup>nd</sup> level therapeutic groups of prescription drugs.

**Results**

Survivors had a 50-60% higher rate of filling prescriptions (table 1) compared with comparisons (RR 1.55, 95% CI 1.52-1.59) and siblings (RR 1.63, 95% CI 1.56-1.67). Stratified by three overall cancer groups (table 1), the highest rate of prescriptions was seen for survivors of CNS tumors (RR 1.96 [95% CI 1.90-2.07]) followed by survivors of hematological malignancies (RR 1.48, 95% CI 1.42-1.54) and non-CNS solid tumors (RR 1.36 [95% CI 1.31-1.41]) compared with the population comparisons.

The relative rates of prescriptions for survivors compared with population comparisons were higher in all 11 ATC 1<sup>st</sup> level and most of the 65 ATC 2<sup>nd</sup> level groups of prescription drugs (figure 1) with the highest rates seen for systematic hormonal preparations and antineoplastic and immunomodulating agents (at the ATC 1<sup>st</sup> level).

The prevalence of polypharmacy increased with age in both survivors and population comparisons (figure 2). At all ages until 70 years the prevalence of polypharmacy increased significantly with age in both survivors and population comparisons during a 120-day period compared with population comparisons, particularly for CNS tumor survivors (see figure 2 and table 2).

**Conclusions**

Our findings emphasize the need for increased awareness for this vulnerable group of childhood cancer survivors at high risk for adverse health outcomes resulting from multiple chronic medications.

confidential


AARHUS  
UNIVERSITY

**Childhood Cancer Research Group (CCA)**

Transcending Boundaries: From Clinical and Medical Labs to Artistic - Evolution of Late Effects Research in Childhood Cancer Survivors

Nermin Ghith (B.Sc., MPH, PhD) on behalf of the research group

**Survivorship and drug AI data lab: Future is here\***

Our research has undergone various phases of transformation, progressing from the utilization of informatics found in clinical records and databases to establishing connections and integrating evidence through extensive national and regional registries.

Today, we stand at the forefront of innovation, harnessing the power of data science and artificial intelligence to revolutionize childhood cancer survivorship. We anticipate biophysiological dimensions of Childhood Cancer by seamlessly integrating data on late effects, treatments, as well as biological and socioeconomic factors.

Contact Nermin Ghith for more information

**PRESENTATION (Publication) NUMBER: P-8**

**P-8**

**1 project**

**INTEGRATED SURVIVORSHIP**

**NOS analysis**

**1 Life after childhood cancer**

**Sc research id in 2010**

**KIWI Knaack Cancer**

**NordForsk**

**5C**

**BARCANCER FONDEN**

**Europen Commission**

**European Union**

**EU FP7**

**PanCare Collaborative**

**novo nordisk foundation**

**ELI-FOU SOLOX**

**Please knock the door: we are very happy to collaborate with our colleagues within and outside the Danish cancer Institute**

**Photos except the one for the shield are downloaded from Unsplash**

Table 1. Rate ratios (RR) of prescriptions and 95% CIs for all childhood cancer survivors and for the sub-groups cohorts of hematological malignancies, CNS tumors or non-CNS solid tumors compared to population and sibling comparisons

	All childhood cancer	Hematological malignancies	CNS tumors	Non-CNS solid tumors
	RR (95% CI)	RR (95% CI)	RR (95% CI)	RR (95% CI)
<b>Survivors of childhood cancer vs population comparisons</b>				
Population comparisons	1.55 (1.52-1.59)	1.48 (1.42-1.54)	1.96 (1.90-2.07)	1.36 (1.33-1.41)
Sex		1 (ref)	1	1
Female	1.76 (1.69-1.82)	1.65 (1.55-1.78)	2.38 (2.23-2.53)	1.48 (1.39-1.56)
Male	1.42 (1.38-1.47)	1.26 (1.29-1.40)	1.79 (1.64-1.88)	1.30 (1.24-1.38)
Age at diagnosis/index date*				
<1	1.64 (1.48-1.82)	1.81 (1.40-2.02)	2.37 (1.93-2.52)	1.45 (1.39-1.52)
1-4	1.64 (1.56-1.72)	1.49 (1.38-1.60)	2.16 (1.95-2.39)	1.51 (1.40-1.64)
5-9	1.62 (1.54-1.73)	1.33 (1.23-1.45)	2.15 (1.93-2.34)	1.35 (1.22-1.51)
10-14	1.59 (1.52-1.67)	1.44 (1.30-1.58)	1.97 (1.81-2.14)	1.37 (1.27-1.49)
15-20	1.45 (1.39-1.51)	1.58 (1.46-1.71)	1.75 (1.61-1.89)	1.30 (1.23-1.38)
Years since diagnosis/index date*				
≤5 years	1.94 (1.87-2.01)	1.79 (1.68-1.93)	2.34 (2.28-2.52)	1.68 (1.59-1.77)
6-10 years	1.61 (1.56-1.66)	1.49 (1.42-1.57)	2.10 (1.98-2.22)	1.49 (1.36-1.50)
10-20 years	1.48 (1.44-1.53)	1.42 (1.34-1.50)	1.88 (1.78-1.98)	1.31 (1.26-1.37)
Test one effect*	<0.0001	>0.0001	>0.0001	>0.0001
<b>Survivors of childhood cancer vs sibling comparisons</b>				
Siblings comparisons	1.63 (1.58-1.67)	1.59 (1.53-1.67)	2.09 (1.97-2.21)	1.39 (1.33-1.45)
Sex		1 (ref)	1	1
Female	1.82 (1.79-1.92)	1.77 (1.63-1.94)	2.36 (2.14-2.68)	1.59 (1.42-1.66)
Male	1.48 (1.42-1.55)	1.44 (1.33-1.55)	1.88 (1.74-2.04)	1.30 (1.22-1.38)
Years since diagnosis/index date*				
≤5 years	1.84 (1.76-1.92)	1.79 (1.66-1.93)	2.36 (2.18-2.55)	1.56 (1.46-1.67)
6-10 years	1.56 (1.50-1.63)	1.48 (1.37-1.59)	1.98 (1.83-2.15)	1.38 (1.29-1.47)
10-20 years	1.61 (1.55-1.68)	1.61 (1.50-1.72)	2.08 (1.99-2.24)	1.37 (1.30-1.45)
Test one effect*	0.0002	>0.0017	>0.0024	

\*Adjusted for country, sex and year group at diagnosis/index date with age as the underlying time scale.

\*Test one effect: Test of whether the exposure effect is the same over the three groups of years since diagnosis/index date.

Photo credit: unsplash website

Figure 1. Rate ratios (RR) and 95% CIs for prescriptions within the 65 ATC 2<sup>nd</sup> level groups of prescription drugs (dispensed more than 100 times among survivors), comparing survivors of childhood cancer with population comparisons

Figure 2. Prevalence and 95% confidence intervals in childhood cancer survivors and population comparisons with polypharmacy

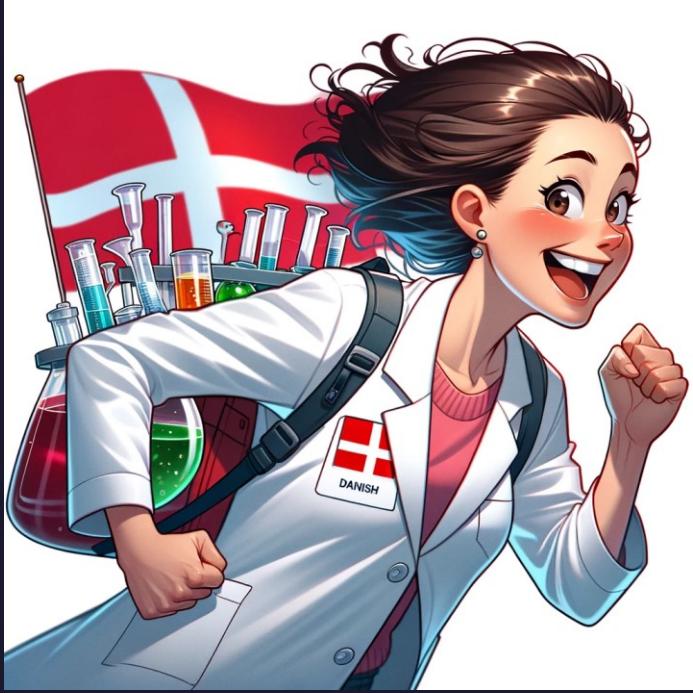
\*Polypharmacy was defined as collecting more than five different prescriptions medications during 120 days below set age.

<sup>a</sup>The P-value is from a chi<sup>2</sup> test comparing the proportion with polypharmacy in the 12 cancer groups and the population comparisons.

28-01-2025

@NerminGhith

48



End of 1<sup>st</sup> symposium  
  
Thank You

Linkedin connect!

