One Data Science Programme

Introduction to Data Science in Pharmaceutical Industry

Dr. Stefanie Müller

Quick Introduction

I am a:

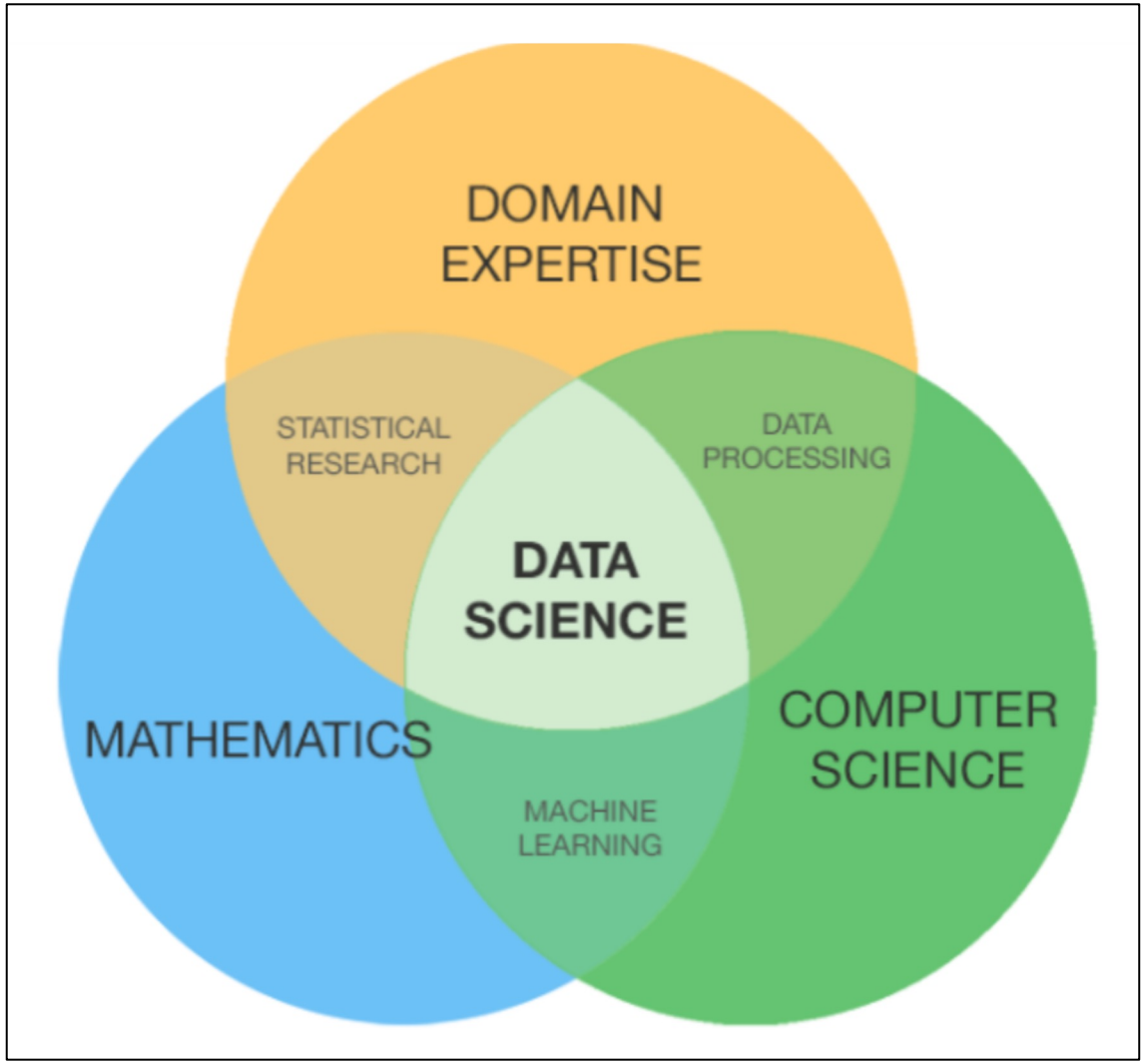
Senior Scientist in the Human Genetics Team in the Research Department of a pharmaceutical company.

I want to talk about:

How me and my colleagues use Data Science to support the development of new medications.

What is Data Science

Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data.



https://towardsdatascience.com/

Questions we try to answer with Data Science

What causes diseases?

Which role does our genetic play in Disease Risk?

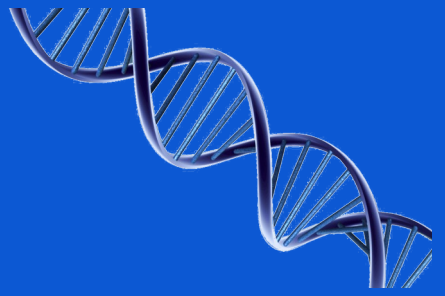
How can we design drugs in a smart way?

What causes diseases?

Lifestyle Factors

Environment

Genetics

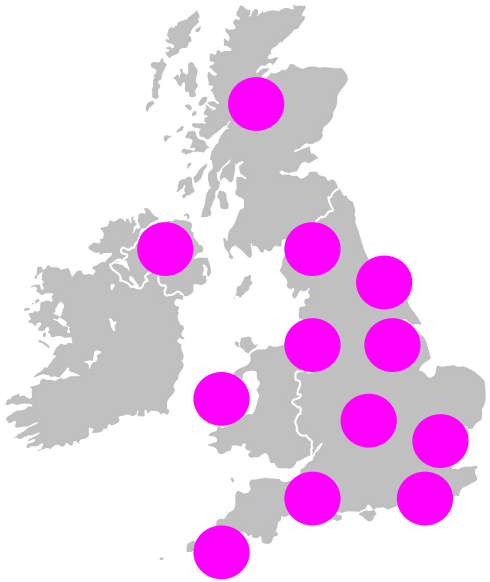


Disease Risk

The higher the specific disease risk the more likely one is to be affected by a disease.

Which role does our genetic play in Disease Risk?

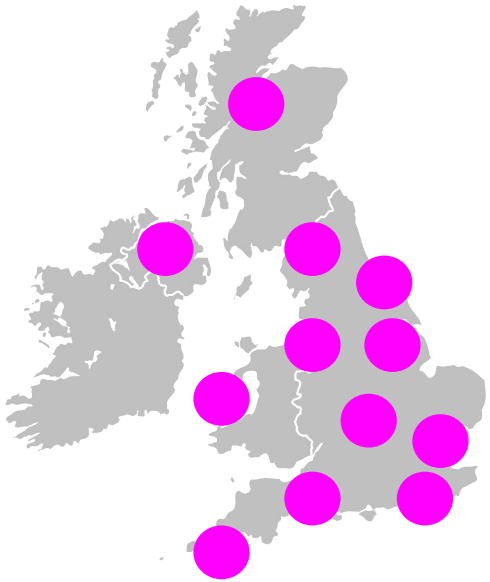
How can we analyse this?



Volunteer
s

Which role does our genetic play in Disease Risk?

How can we analyse this?

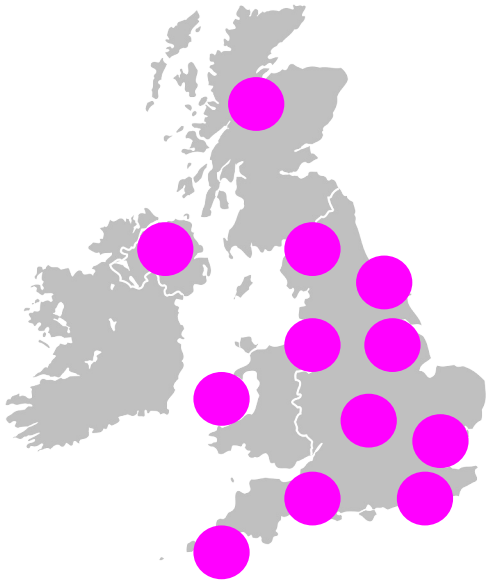


Volunteers

Medical History

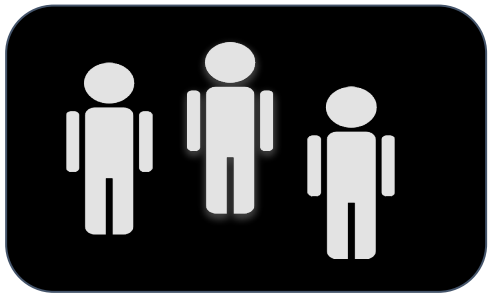
Which role does our genetic play in Disease Risk?

How can we analyse this?



Volunteers

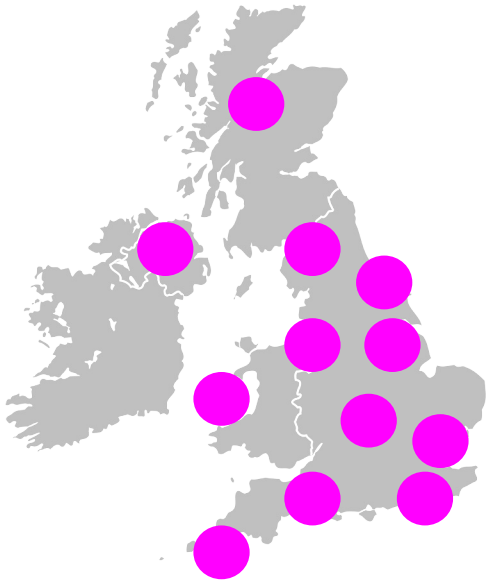
Medical History



Find out who is affected

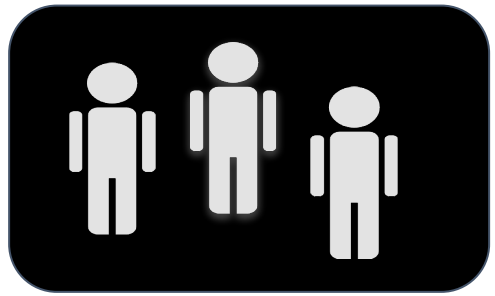
Which role does our genetic play in Disease Risk?

How can we analyse this?

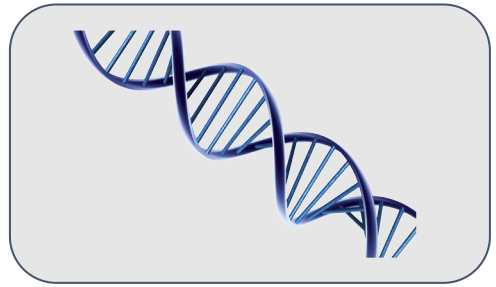


Volunteers

Medical History



Find out who is affected



Compare DNA

Which role does our genetic play in Disease Risk?

How can we analyse this?

Such data collections exist and one of the most important ones is the UK Biobank

Volunteers

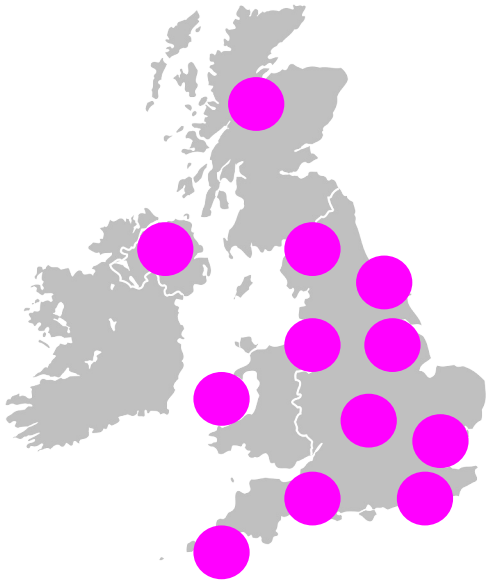
Medical History

Find out who is affected

Compare DNA

Which role does our genetic play in Disease Risk?

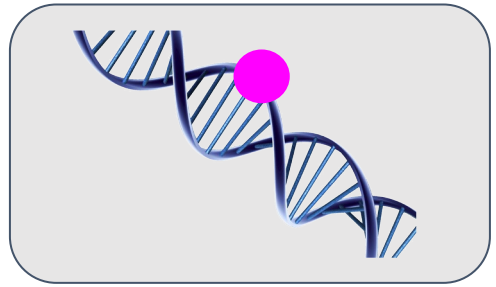
How can we analyse this?

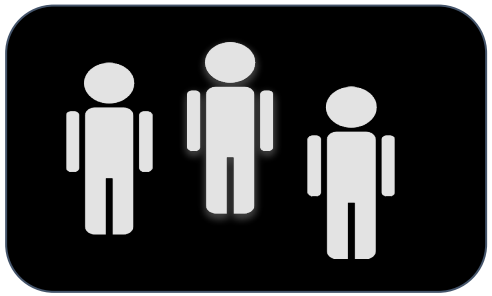


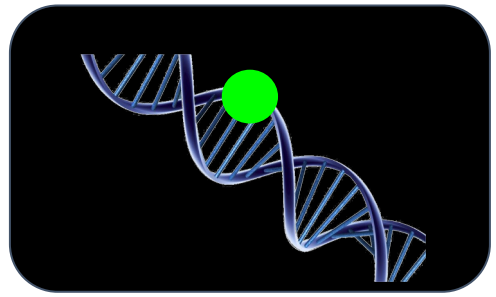
Volunteers

Medical History

Find out who is affected





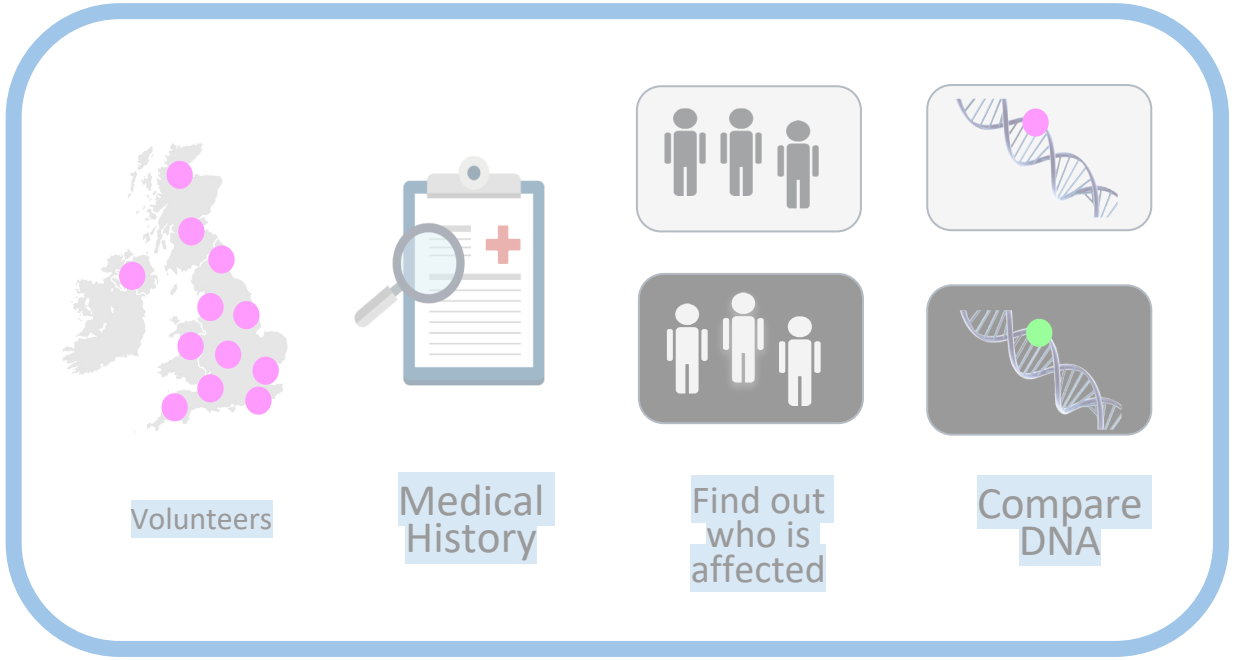


Compare DNA

Which role does our genetic play in Disease Risk?

How was Data Science used?

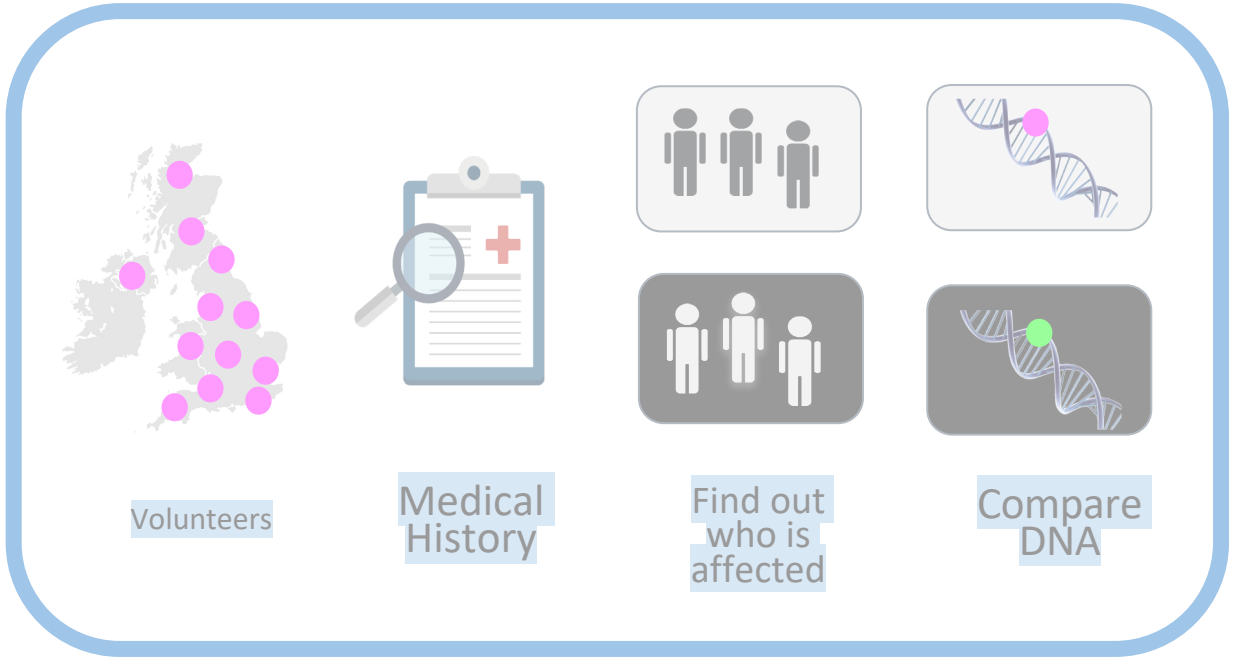
- Combined data from different data sources
- Cleaned data of missing or faulty values
- Used statistical models to find differences between groups
- Visualised results

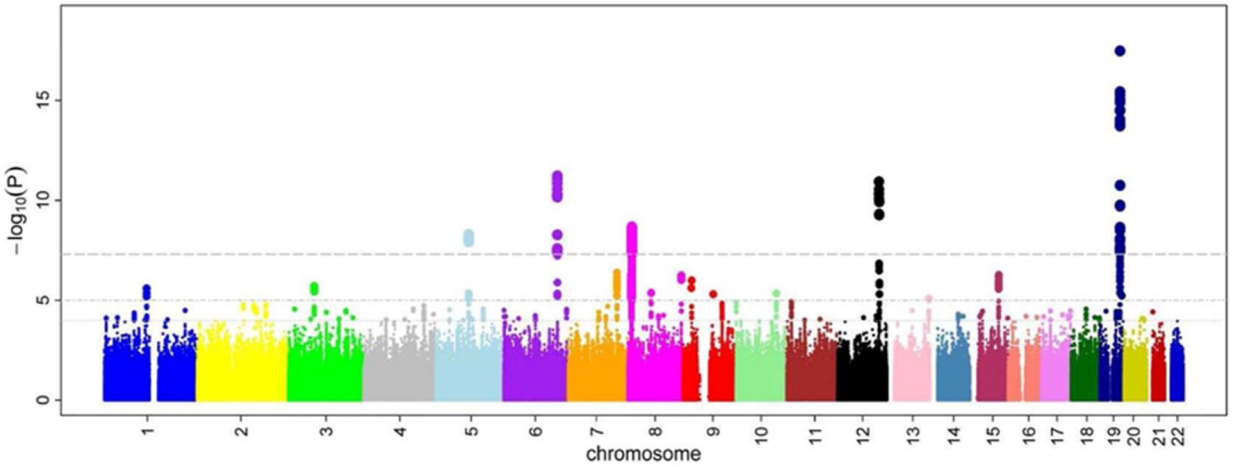


Which role does our genetic play in Disease Risk?

How was Data Science used?

- Combined data from different data sources
- Cleaned data of missing or faulty values
- Used statistical models to find differences between groups
- Visualised results

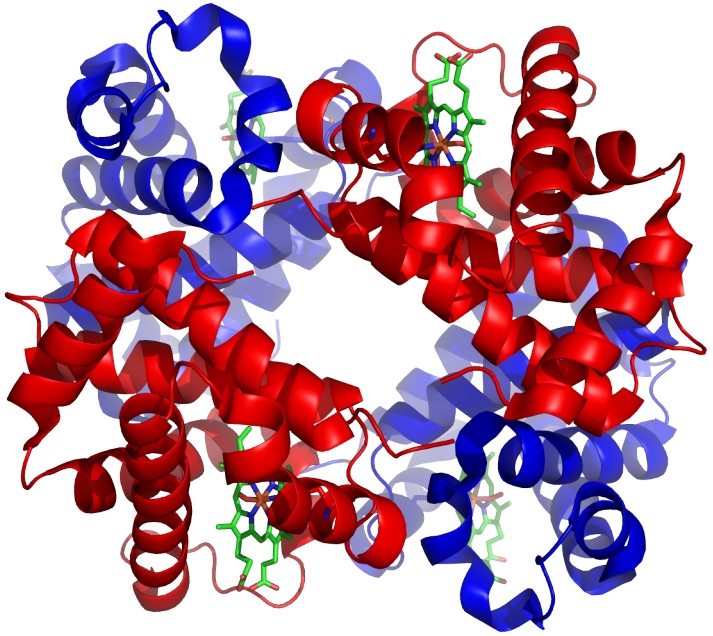




How can we design drugs in a smart way?

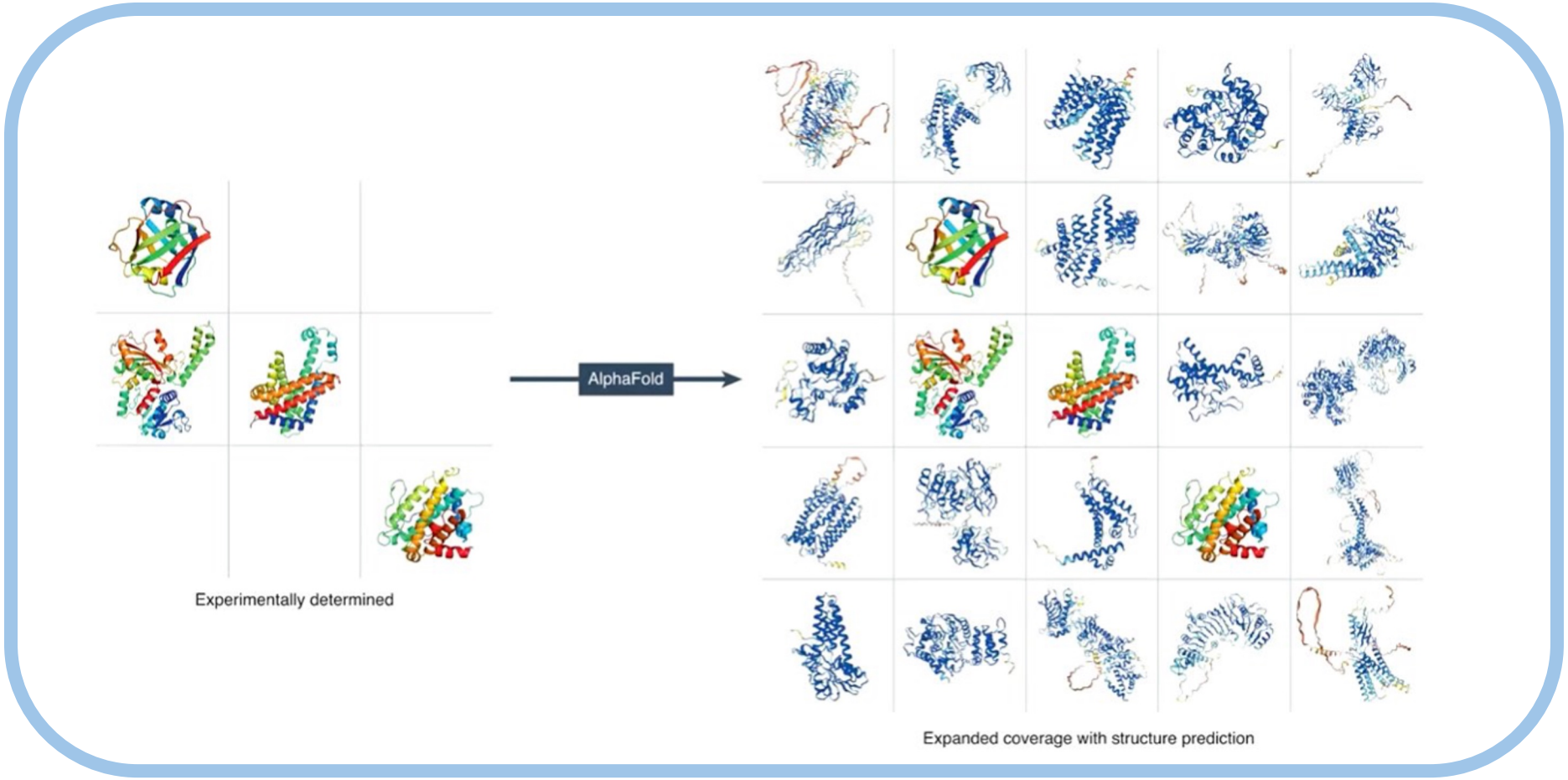
Genetic differences can help us understand which proteins are important in a disease

- Proteins are involved in every biological process
- A common medication approach can be to block a protein from its function
- In these cases protein and drug oftentimes fit together like a lock and its key
- Knowing the 3D structure of proteins can help finding the correct keys
- It can take years to find correct structure via experiments



hemoglobin

How can we design drugs in a smart way?

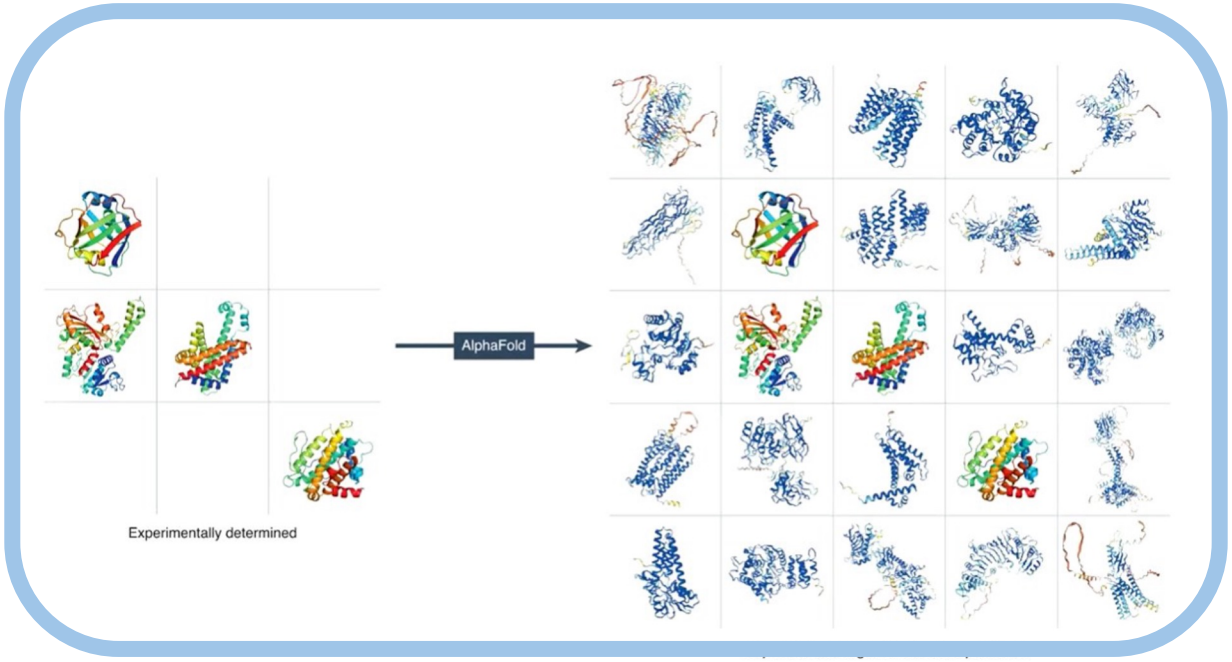


Alphafold is an AI algorithm which learns from patterns from already solved protein structures and can in this way lead to completely new 3D structures for formerly unknown proteins

How can we design drugs in a smart way?

How was Data Science used?

- Artificial intelligence and machine was used to learn complex patterns and rules from existing data
- These learnings, called models, are used to infer new data
- This is a pattern we expect to see much more often in Life Science and Biology in general



Further Information

UK Biobank:

- https://www.youtube.com/watch?v=66moI1ZHMYs
- https://en.wikipedia.org/wiki/UK_Biobank

Alphafold:

- https://www.youtube.com/watch?v=KpedmJdrTpY
- https://www.deepmind.com/research/highlighted-research/alphafold