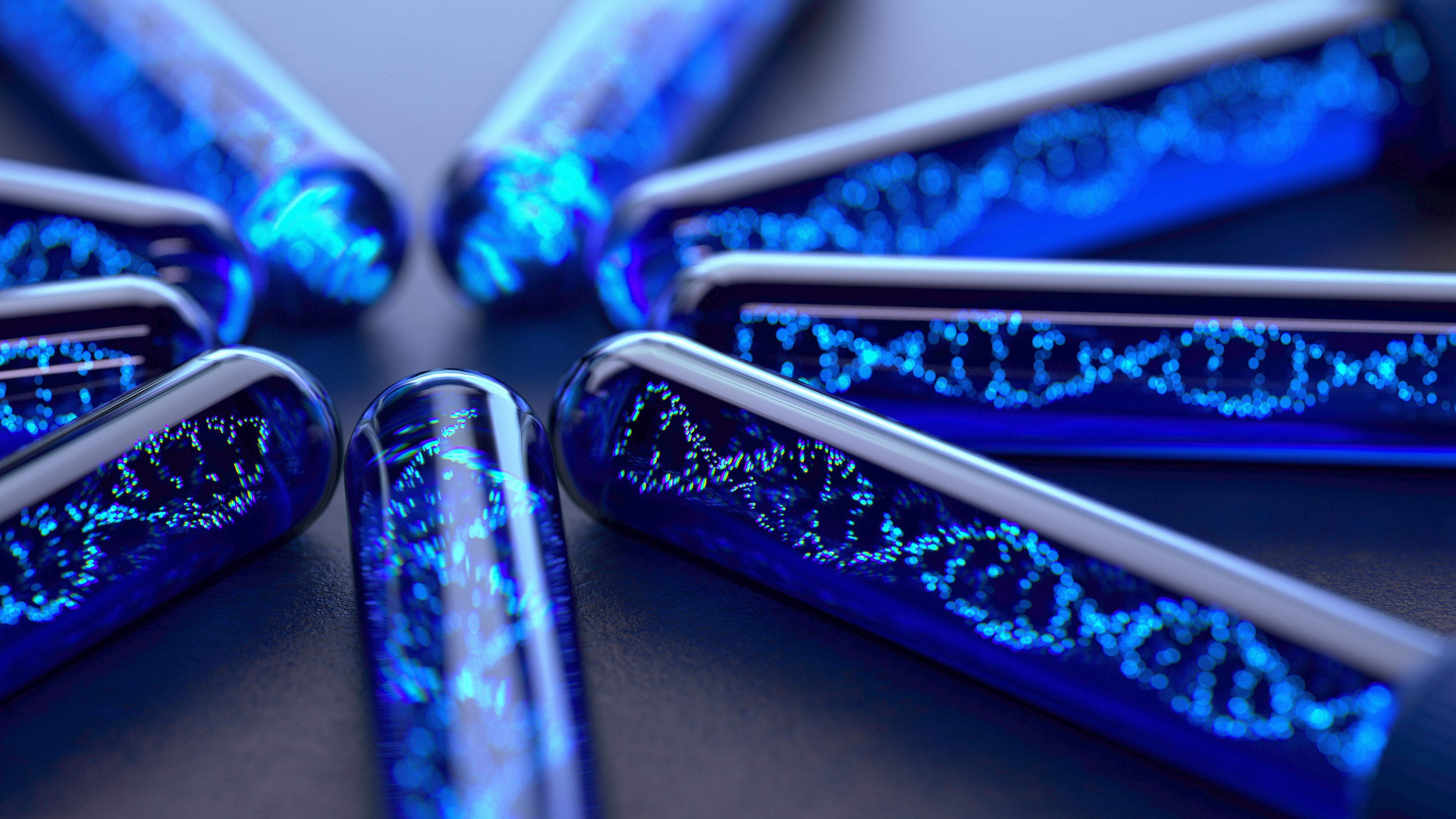


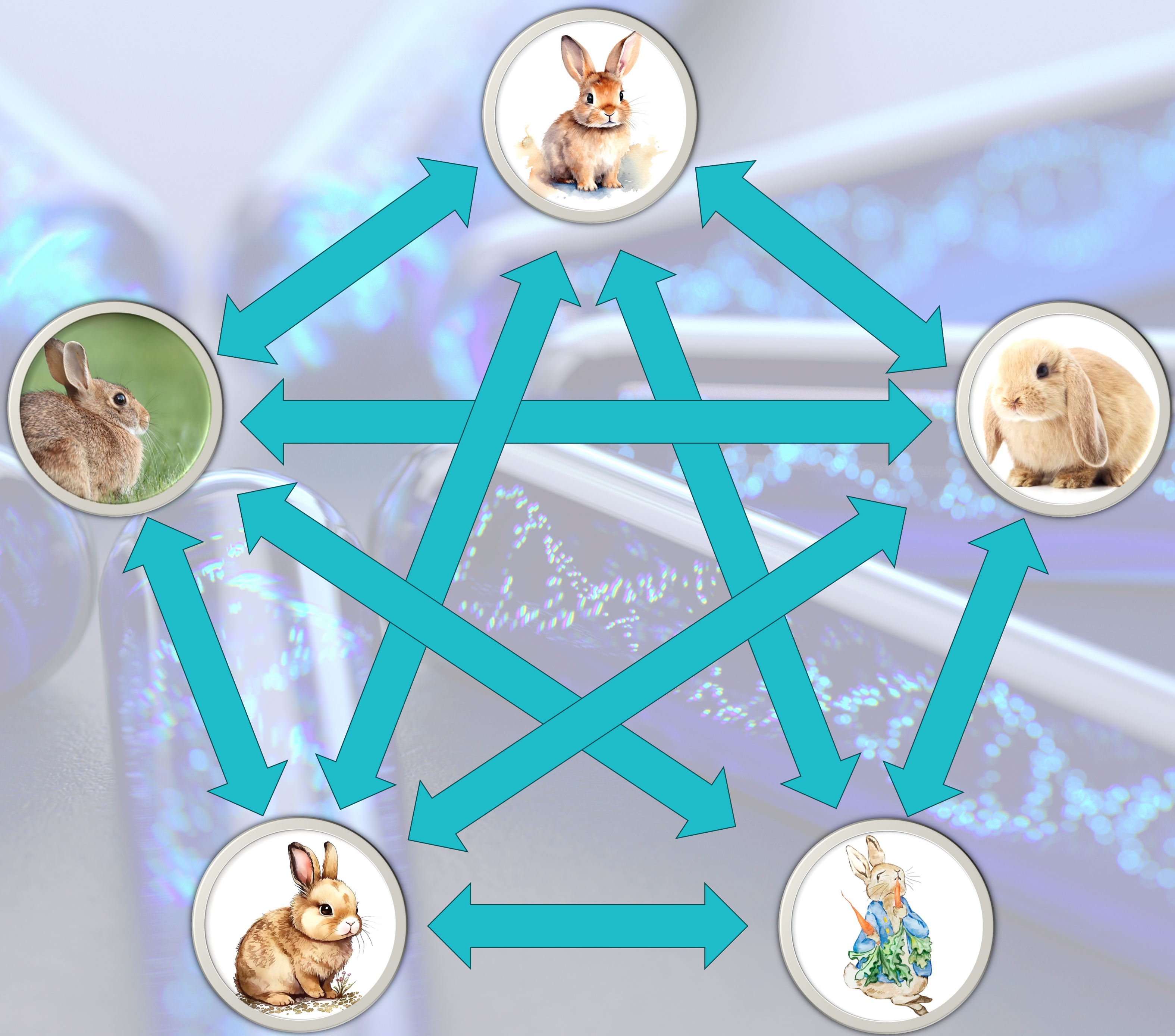
Ben Below - Softwire Chromosomes in the cloud

Building a scalable genomic
matching algorithm











Your immune system can identify you



**Your immune system can identify you,
and you're unique!**



**ANTHONY
NOLAN**



ANTHONY

Atlas Public

Edit Pins Watch 9 Fork 5 Starred 9

master 20 Branches 25 Tags

Go to file Add file Code

pchenery feature: #1306: CBT-1822: Add query to find patients and do... 489c36e · 4 days ago 4,901 Commits

.github/ISSUE_TEMPLATE	chore: New enhancement request issue template.	2 years ago
ArchitecturalDecisionRecord	chore: #1295: Debug client packages now versioned in st...	last month
ArchitectureDiagrams	docs: #652 Add overall architecture diagram for Atlas	3 years ago
Atlas.Client.Models	chore: #1285: Update Newton.Json package to latest ver...	last month
Atlas.Common.Public.Models	feature: #1294: Added ability to process MV4 homework	2 weeks ago
Atlas.Common.Test	refactor: #1241: Allow for multiple sb connection strings i...	3 months ago
Atlas.Common	feature: #1294: Added ability to process MV4 homework	2 weeks ago
Atlas.Debug.Client.Models	feature: #1294: Added ability to process MV4 homework	2 weeks ago
Atlas.Debug.Client	chore: #1285: Update Newton.Json package to latest ver...	last month

About

A free & open-source Donor Search Algorithm Service

open-source search-algorithm match-prediction stem-cell-transplant hla-matching wmda anthony-nolan donor-search

Readme GPL-3.0 license Activity Custom properties 9 stars 9 watching 5 forks Report repository

Languages

C#	95.1%	HCL	2.5%
Gherkin	1.8%	JavaScript	0.4%
TSQL	0.2%	PowerShell	0.0%



Suitability for stem cell transplantation is determined primarily by 6 HLA genes

We may only know that a donor has one of 500 different options for each gene

No Ambiguity

A*01:01

Some Ambiguity

A*01:AB

Large Ambiguity

A*01:XX

A*01:01

OR

A*01:02

A*01:01

OR

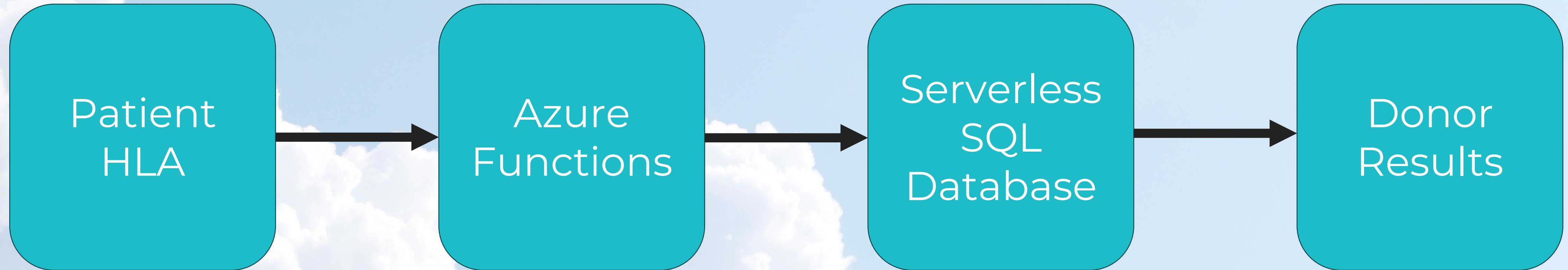
A*01:02

OR 453 MORE...

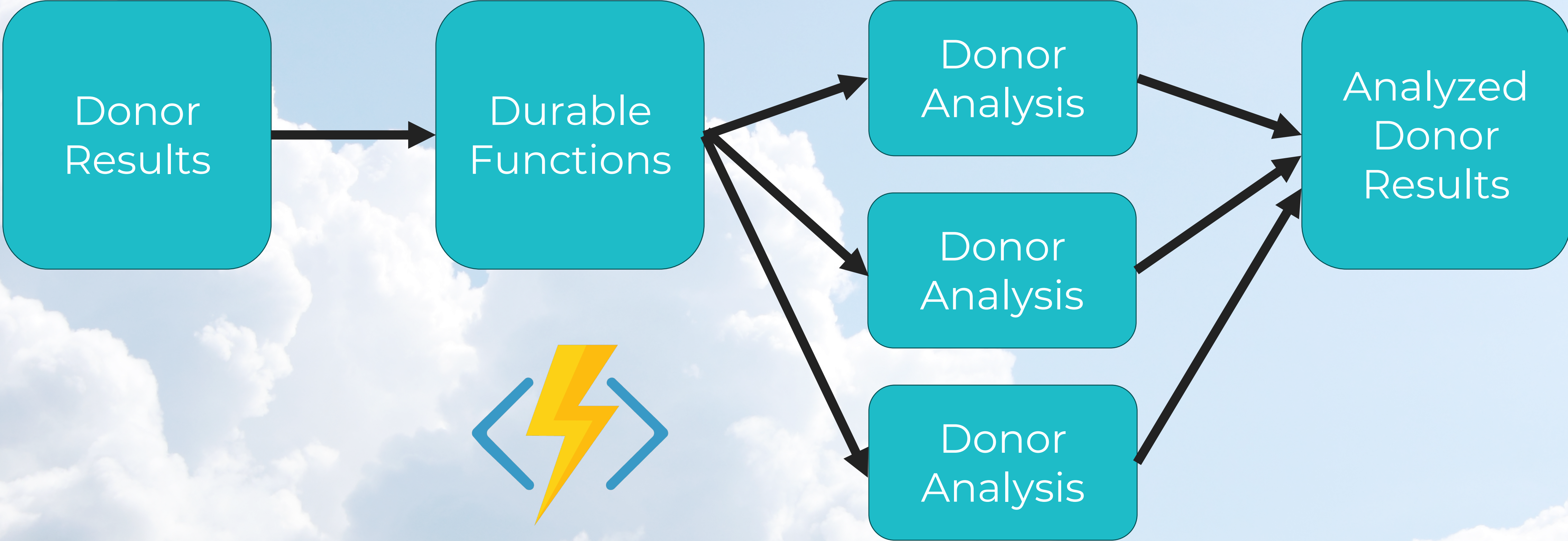
A scientist wearing a white protective suit, mask, and goggles is focused on a microscope in a laboratory. The scene is lit with a cool blue tone. In the background, another person in a white lab coat is visible, working at a similar workstation. The overall atmosphere is one of precision and scientific inquiry.

We worked closely with Anthony Nolan's scientists to build a scalable algorithm

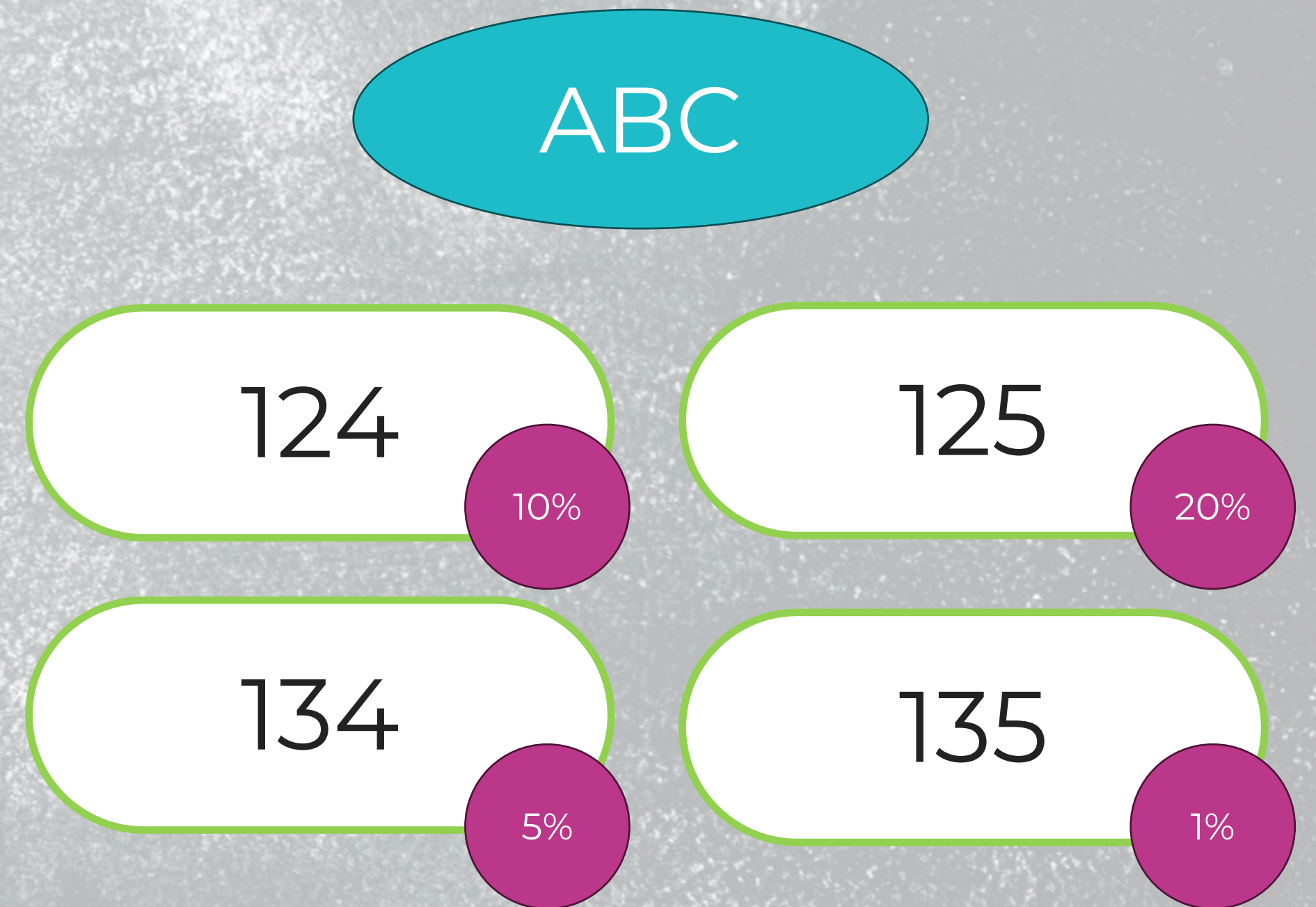
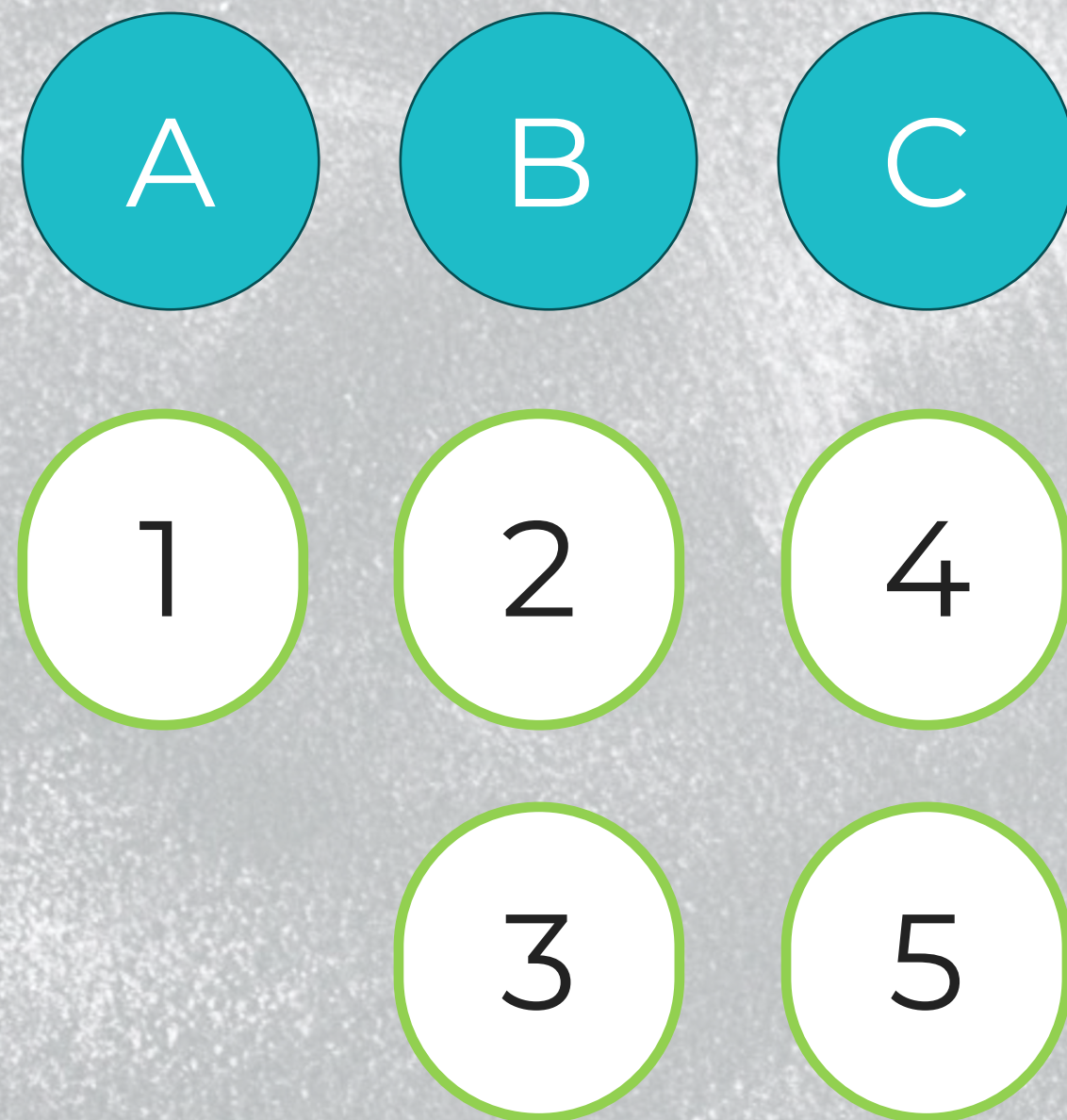
Matching Algorithm

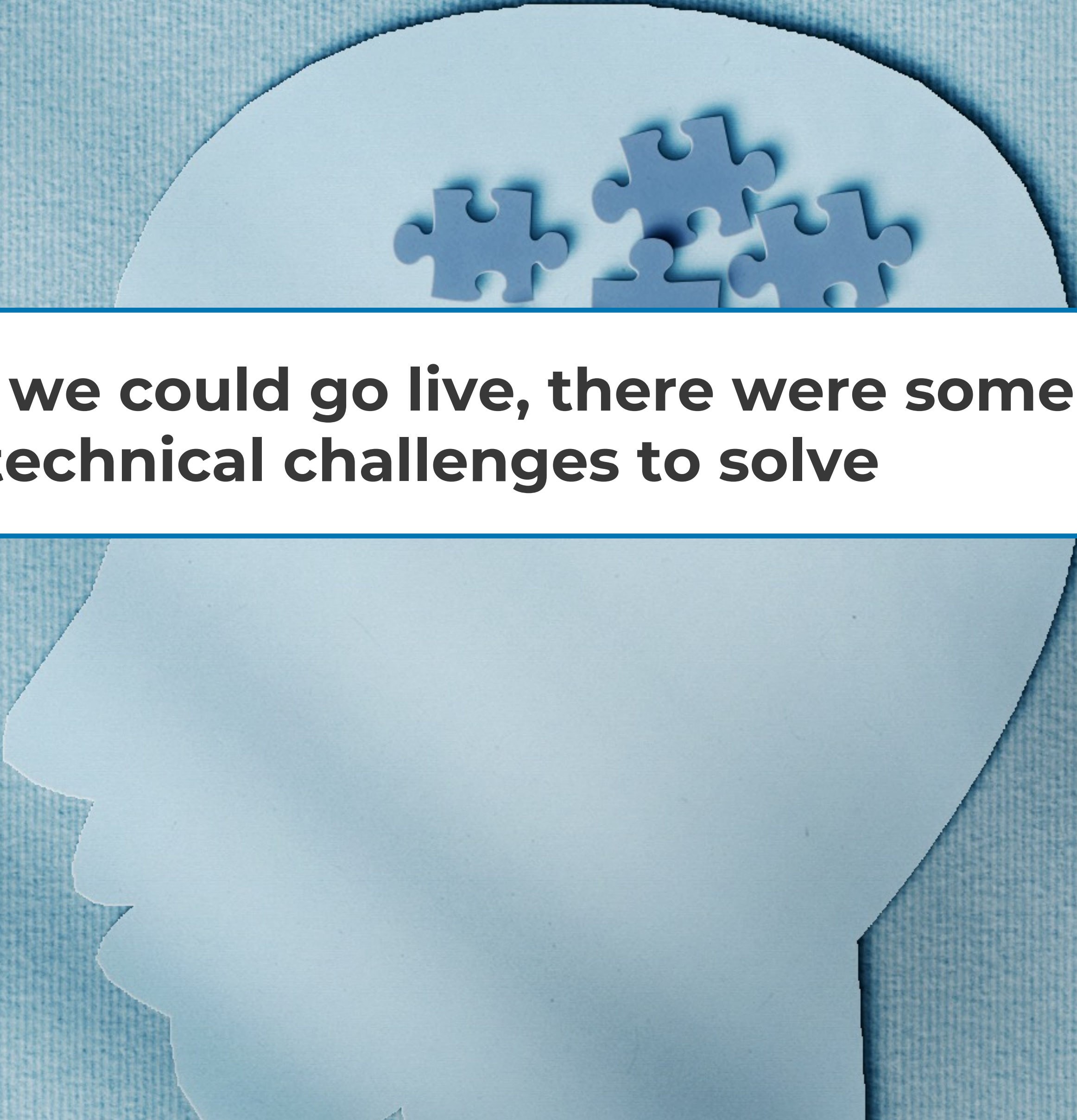


Match Prediction Algorithm



Donor Analysis

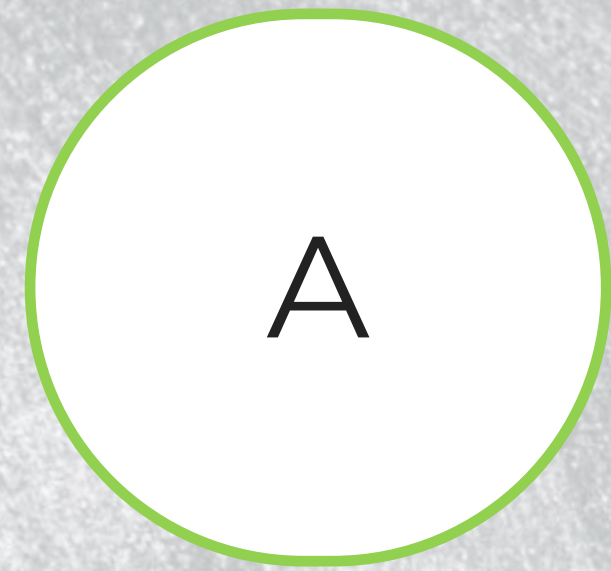


A light blue silhouette of a human head in profile, facing left. Inside the head, three puzzle pieces are arranged in a small cluster. The background is a textured, light blue surface.

**Before we could go live, there were some
technical challenges to solve**



The ambiguity problem returns – and we run completely out of memory



A

B

C

DRB1

DQB1

Gigabyte = 10^9

Terabyte = 10^{12}

Zettabyte = 10^{21}

Yottabyte = 10^{24}



No computer on Earth could hold all possible genetic combinations



Only expand to options that exist in the reference data, not every possible one!

A night scene featuring palm trees in the foreground and several bright, golden-yellow fireworks exploding in the dark sky. The fireworks are in various stages of explosion, with some showing long, thin trails and others forming large, starburst patterns. The palm trees are silhouetted against the dark background, with some fronds catching the light from the fireworks.

**Success! The out of memory errors are
vanquished!**



**Our calculations will take hours.. for
each donor result**



**Our calculations will take hours.. for
each donor result**

0: 2%
1: 10%
2: 7%

M

0%

01:ABGEP

01:ABGEP

A

100%

01:NRAR

35:01:01

A

100%

01:NRAR

35:01:01

L

0%

04:03

11:01

P

10%

03:02:01

03:01:01

P

11:01:01G

04:03:01G

7450 
IT-Milano

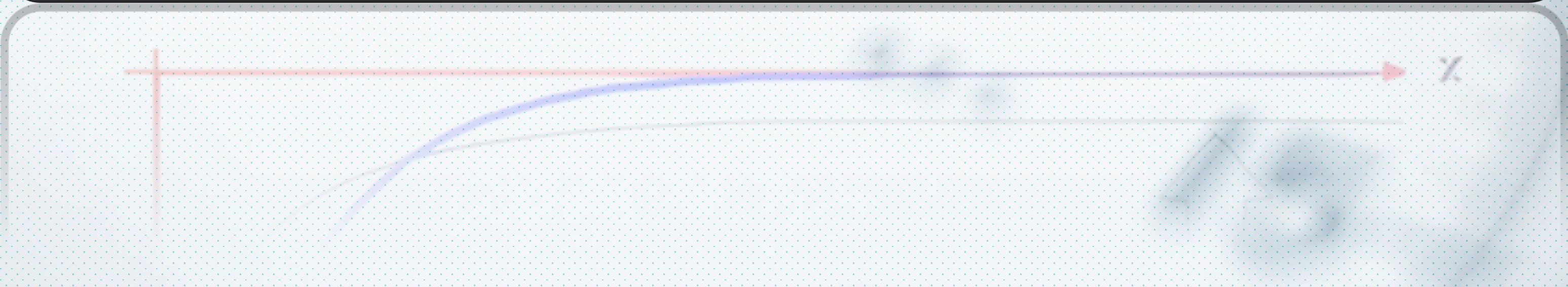
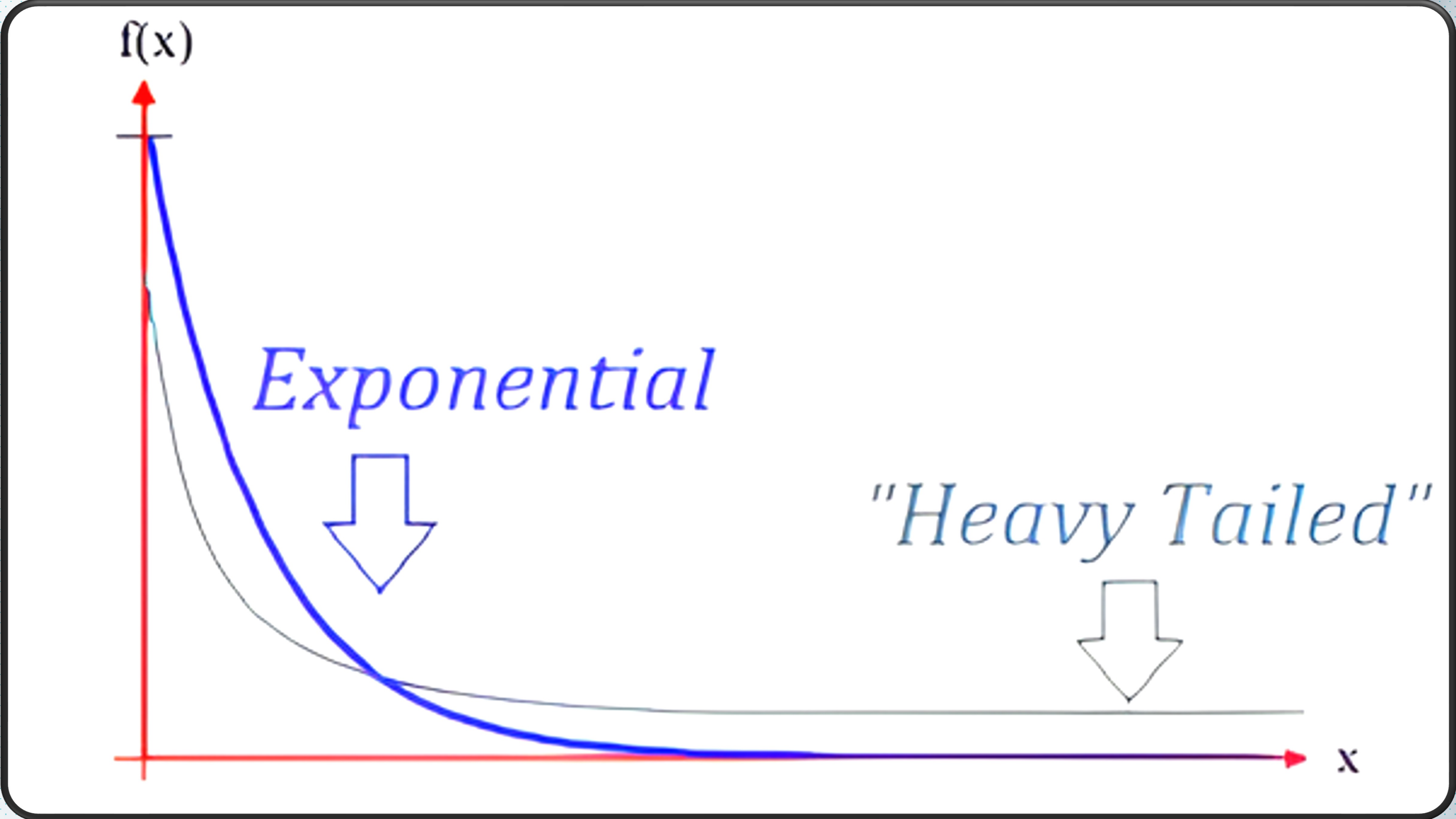
Female

25

O+



The sets of genes involved differ in likelihood by 10 orders of magnitude





**We don't need to run all the calculations
– most of them are noise!**

Takeaways



Flip the problem around



Takeaways



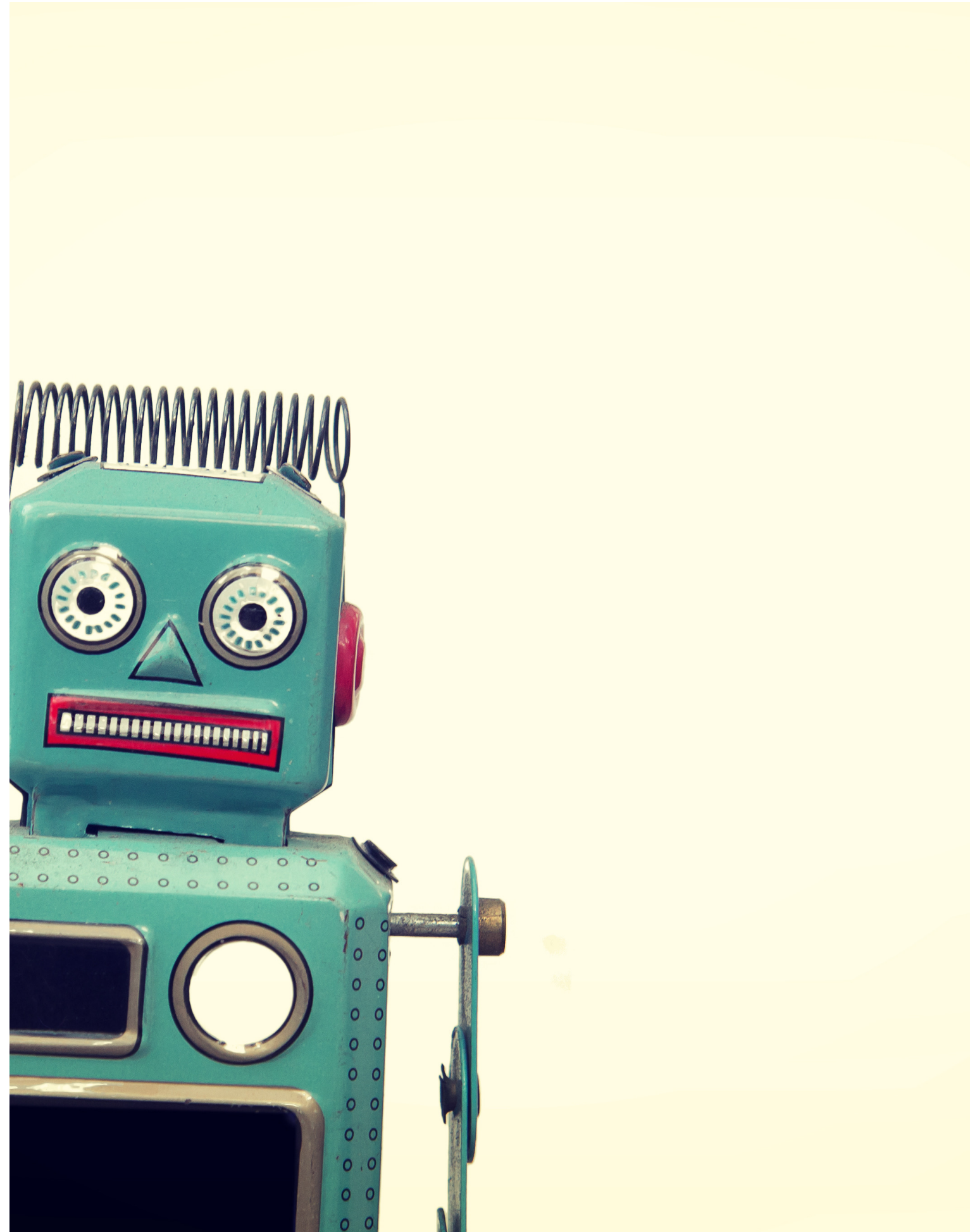
Flip the problem around



Be prepared to trade precision for speed



Takeaways



- 1 Flip the problem around
- 2 Be prepared to trade precision for speed
- 3 The solution often isn't a technical one

Takeaways



1

Understand your domain

2

3

4



Takeaways



Understand your domain



Ask the right questions



Takeaways



Understand your domain



Ask the right questions



Seek out and challenge NFRs



Takeaways



- 1 Understand your domain
- 2 Ask the right questions
- 3 Seek out and challenge NFRs
- 4 Join the register!



Ben Below

Technical Principal – Softwire

Email: benjamin.below@softwire.com

LinkedIn: www.linkedin.com/in/ben-below/



Thank you