

Jörg Menche

Unlocking the secrets of biology through art, imaging and data science

3rd International Danube Symposium| 21.09.2023

 joerg@menchelab.com

 www.menchelab.com

 [@menchelab](https://twitter.com/menchelab)

Jörg Menche

Unlocking the secrets of biology through art, imaging and data science

Perhaps

Hmmm

No, sorry

Yes?

3rd International Danube Symposium | 21.09.2023

✉ joerg@menchelab.com

🌐 www.menchelab.com

🐦 @menchelab

Julia Guthrie

Exploring inflammation in 3D

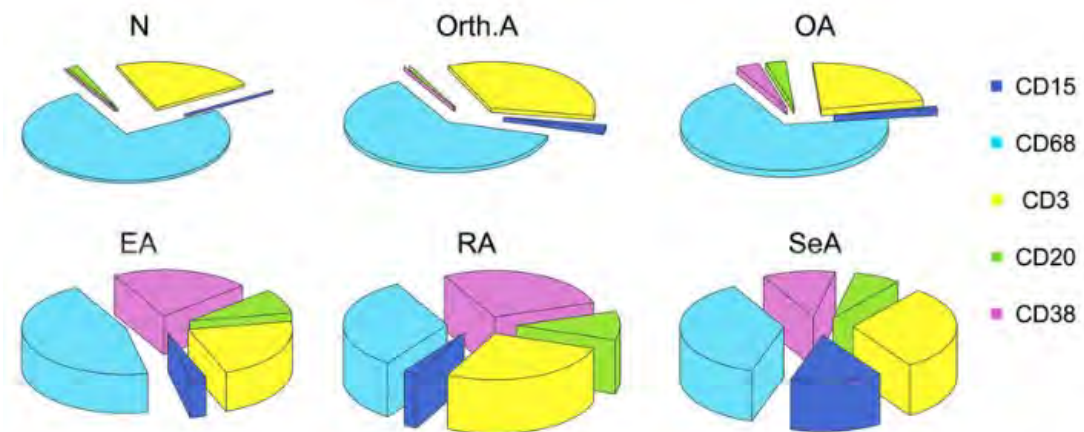
3rd International Danube Symposium | 21.09.2023

 julia@menchelab.com

 www.menchelab.com

 [@menchelab](https://twitter.com/menchelab)

Inflammation in 3D



Three-dimensional pie charts illustrating the inflammatory gestalts of the arthropathies and normal synovium. The pie sections correspond to the percentages of specific cell types in the inflammatory infiltrates. The height of each pie corresponds to the trimmed mean of the sum of the cells expressing any of the five markers ("total number of inflammatory cells") and is used to express the overall degree of inflammation. doi: 10.1371/journal.pone.0072494.g002

RENCHÉ LAB



RANDOM WALK GO!

RESISTANT PROBABILITY: 0.54

█	█
█	█
█	█
█	█
█	█
█	█
█	█
█	█
█	█
█	█

INSPECTOR

SMR3A
submaxillary gland androgen regu
3A
2 Neighbors



The basic building blocks of life

Initial sequencing and analysis of the human genome | Nature

nature.com/articles/350...

Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium

Nature 409, 860–921 (2001) | [Cite this article](#)

270k Accesses | 16490 Citations | 1315 Altmetric | [Metrics](#)

A [Corrigendum](#) to this article was published on 01 August 2001

An [Erratum](#) to this article was published on 01 June 2001

Abstract

The human genome holds an extraordinary trove of information about human development, physiology, medicine and evolution. Here we report the results of an international collaboration to produce and make freely available a draft sequence of the human genome. We also present an initial analysis of the data, describing some of the insights that can be gleaned from the sequence.

Main

The rediscovery of Mendel's laws of heredity in the opening weeks of the 20th century^{1,2,3} sparked a scientific quest to understand the nature and content of genetic information that has propelled biology for the last hundred years. The scientific progress made falls naturally into four main phases, corresponding roughly to the four quarters of the century. The first established the cellular basis of heredity: the chromosomes. The second defined the molecular basis of heredity: the DNA double helix. The third unlocked the informational basis of heredity, with the discovery of the biological

You have full access to this article via Universitaet Wien Bibliotheks- und Archivwesen

[Download PDF](#)

Associated Content

Collection

40 years of Sanger sequencing

Collection

The 1000 Genomes Project

Sections | **Figures** | References

[Abstract](#)

[Main](#)

[Background to the Human Genome Project](#)

[Strategic issues](#)

[Generating the draft genome sequence](#)

[Broad genomic landscape](#)

[Repeat content of the human genome](#)

[Simple sequence repeats](#)

[Gene content of the human genome](#)

[Protein-coding genes](#)

[Comparative proteome analysis](#)

[Conclusion](#)

Highly accurate protein structure prediction with AlphaFold | Nature

nature.com/articles/s415...

Highly accurate protein structure prediction with AlphaFold

John Jumper, Richard Evans, ... Demis Hassabis | [+ Show authors](#)

Nature 596, 583–589 (2021) | [Cite this article](#)

685k Accesses | 2020 Citations | 3072 Altmetric | [Metrics](#)

Abstract

Proteins are essential to life, and understanding their structure can facilitate a mechanistic understanding of their function. Through an enormous experimental effort^{1,2,3,4}, the structures of around 100,000 unique proteins have been determined⁵, but this represents a small fraction of the billions of known protein sequences^{6,7}. Structural coverage is bottlenecked by the months to years of painstaking effort required to determine a single protein structure. Accurate computational approaches are needed to address this gap and to enable large-scale structural bioinformatics. Predicting the three-dimensional structure that a protein will adopt based solely on its amino acid sequence—the structure prediction component of the ‘protein folding problem’⁸—has been an important open research problem for more than 50 years⁹. Despite recent progress^{10,11,12,13,14}, existing methods fall far short of atomic accuracy, especially when no homologous structure is available. Here we provide the first computational method that can regularly predict protein structures with atomic accuracy even in cases in which no similar structure is known. We validated an entirely redesigned version of our neural network-based model, AlphaFold, in the challenging 14th Critical Assessment of protein Structure Prediction (CASP14)¹⁵, demonstrating accuracy competitive with experimental structures in a majority of cases and greatly outperforming other methods. Underpinning the latest version of AlphaFold is a novel machine learning approach that incorporates physical and biological knowledge about protein structure, leveraging

[Download PDF](#)

Associated Content

Nature Outlook

Robotics and artificial intelligence

Focus

Method of the Year 2021: Protein structure prediction

Highly accurate protein structure prediction for the human proteome

Kathryn Tunyasuvunakool, Jonas Adler, ... Demis Hassabis

Nature [Article](#) [Open Access](#) 22 Jul 2021

Protein-structure prediction revolutionized

Mohammed AlQuraishi

Nature [News & Views](#) 23 Aug 2021

Sections | **Figures** | References

[Abstract](#)

[Main](#)

[The AlphaFold network](#)

[Evoformer](#)

[End-to-end structure prediction](#)

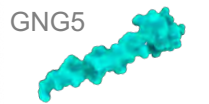
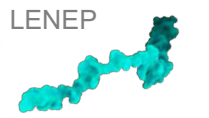
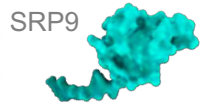
[Training with labelled and unlabelled data](#)

[Interpreting the neural network](#)

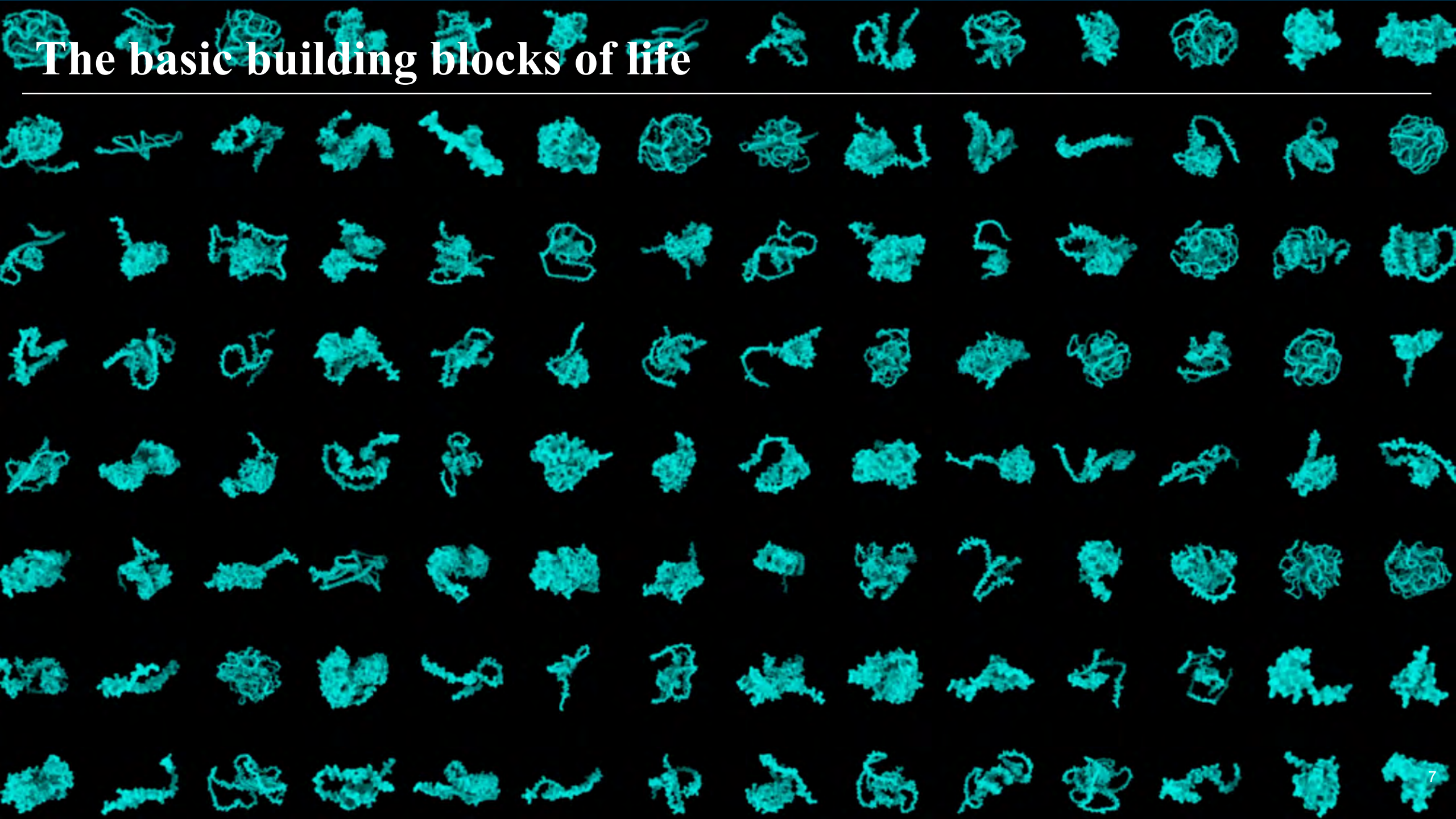
23 chromosomes

3 billion base pairs

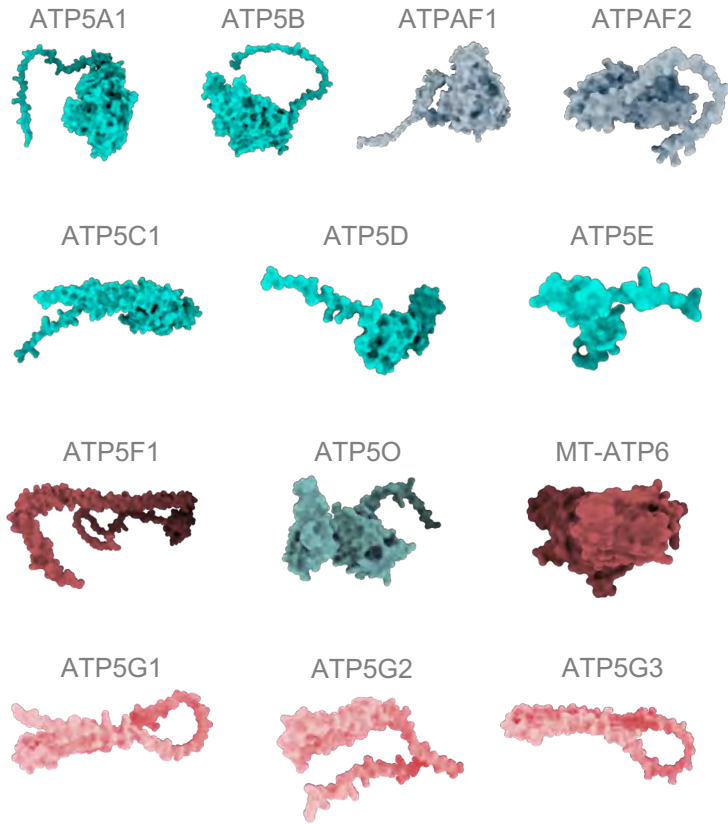
20,000 genes



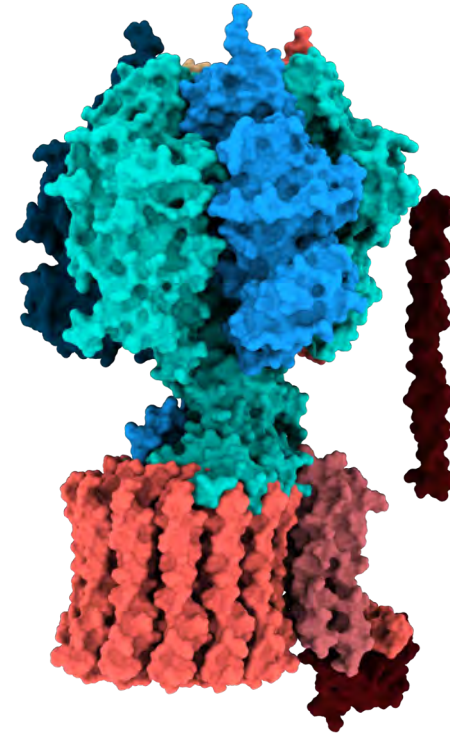
The basic building blocks of life



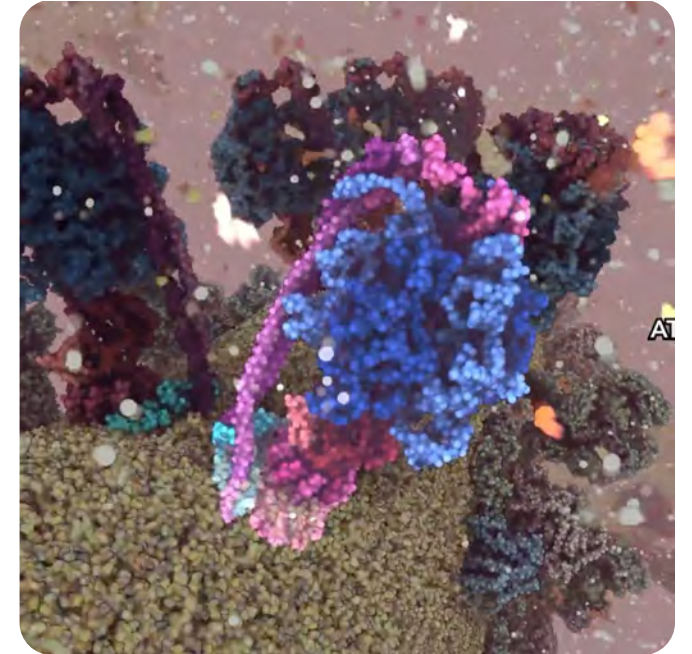
Putting the pieces together



The components of the ATP synthase

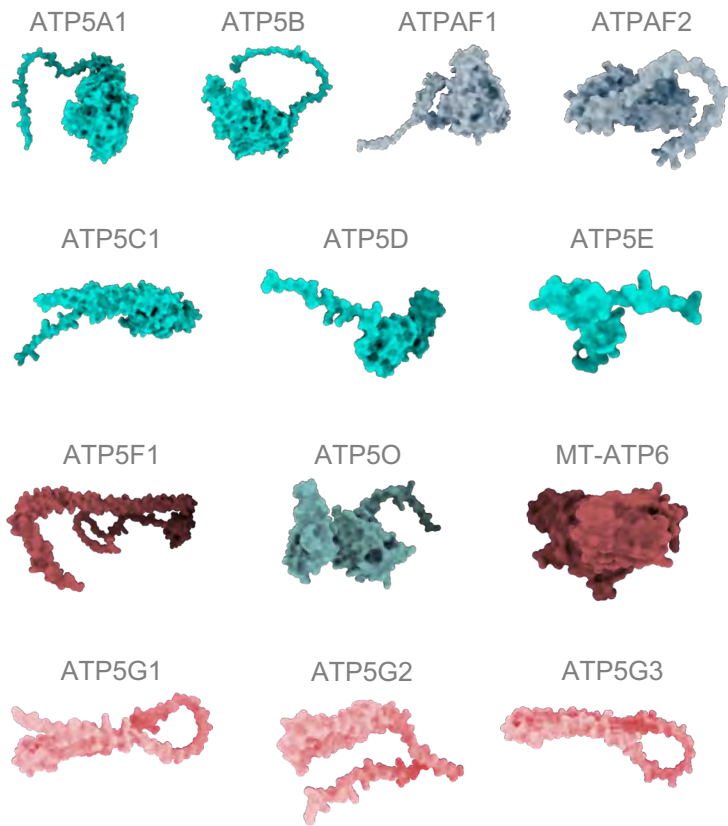


The assembled complex

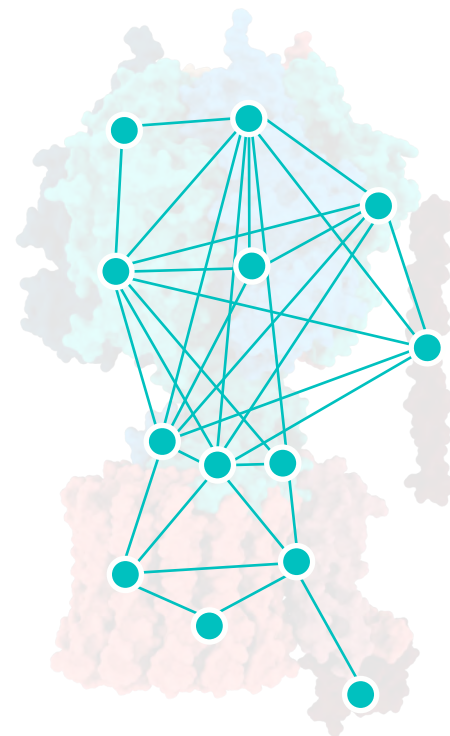


The complex in action

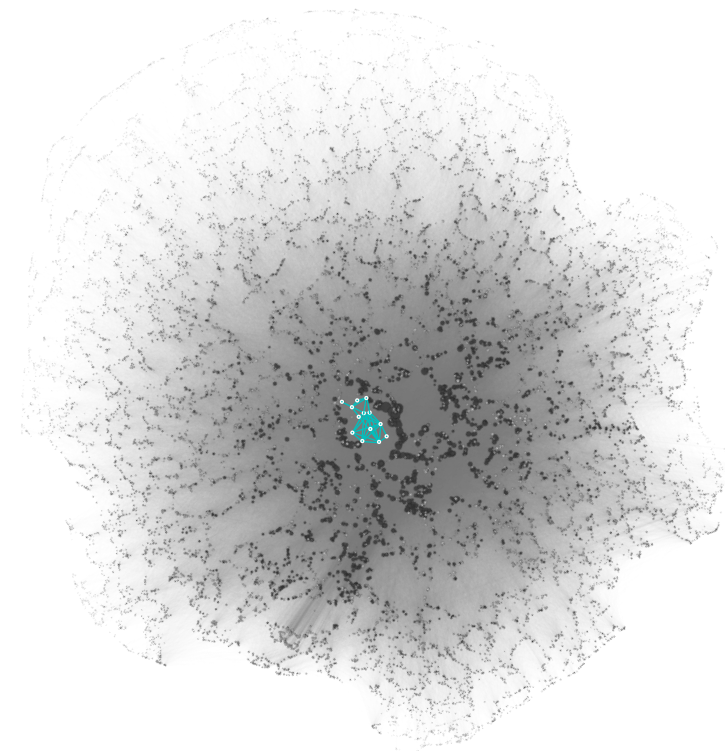
The interactome provides a map for biological functions



The components of the ATP synthase



The assembled complex



The complex within the interactome

Networks as maps



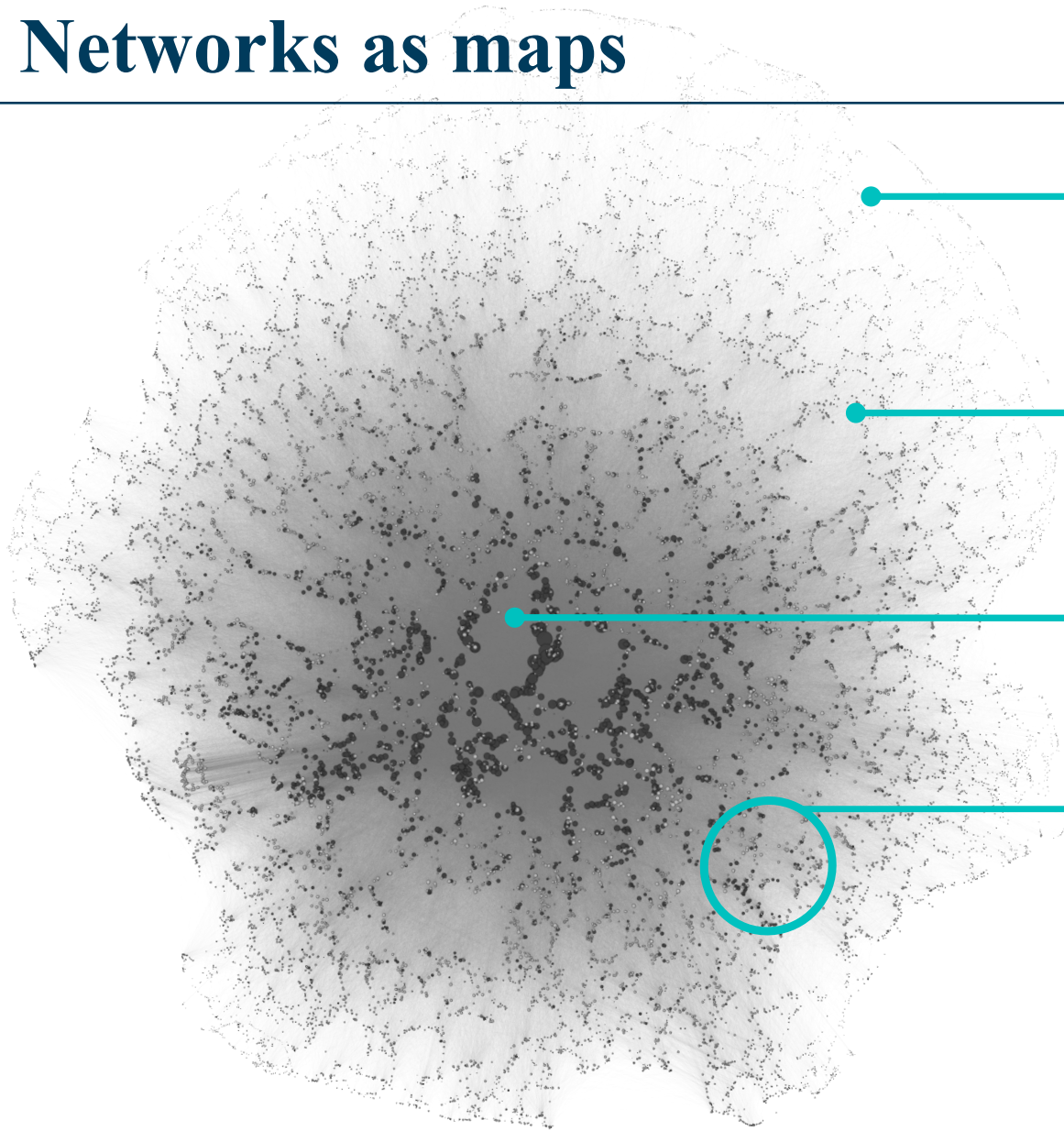
Distance reflects independence

Global structure reflects historical development and environmental constraints

Location matters

Neighborhoods reflect relatedness

Networks as maps



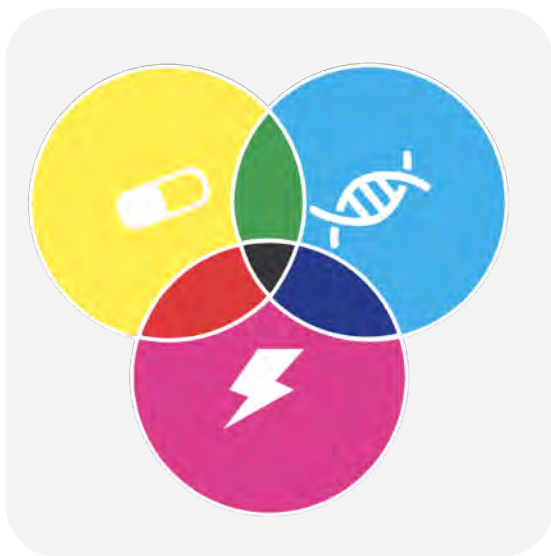
Distance reflects independence

Global structure reflects historical development and environmental constraints

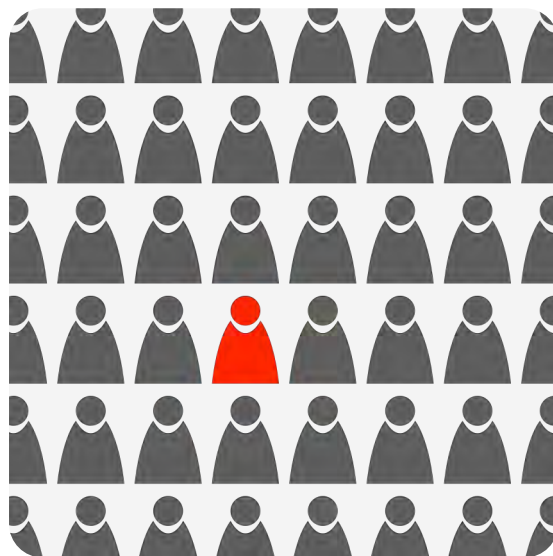
Location matters

Neighborhoods reflect relatedness

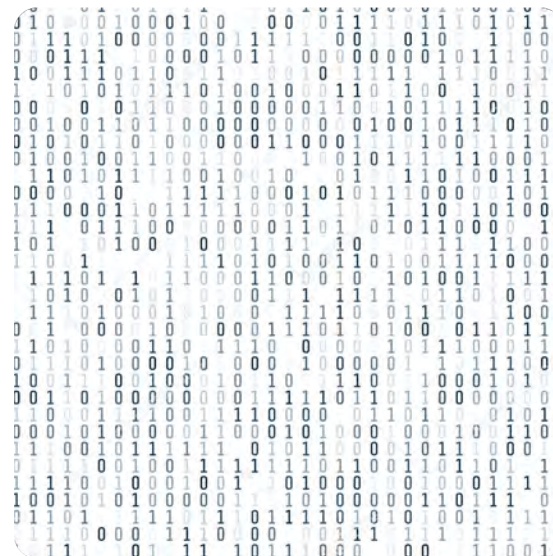
Where do we go from here?



Fundamental organizational principles in biology



Concrete impact on medical diagnosis and treatment

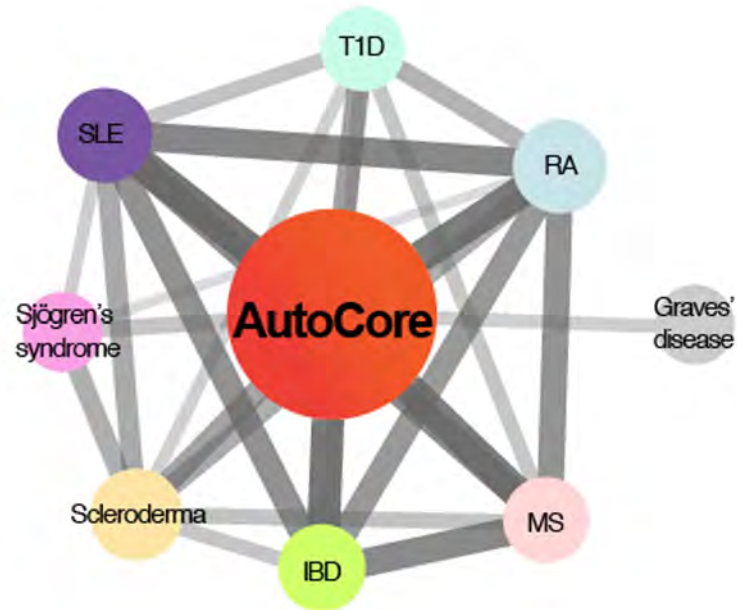


Open up new fields for network science



Pioneer new tools for data exploration

What you will see today

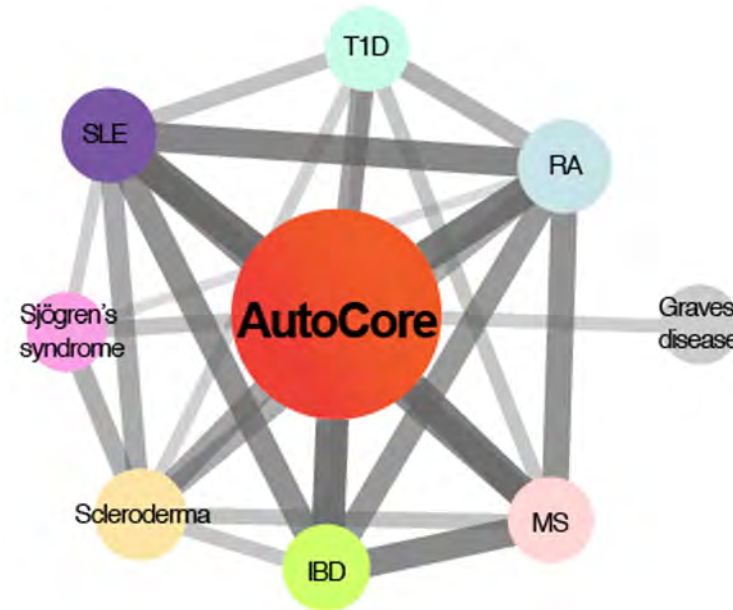


Networks of autoimmunity and autoinflammation



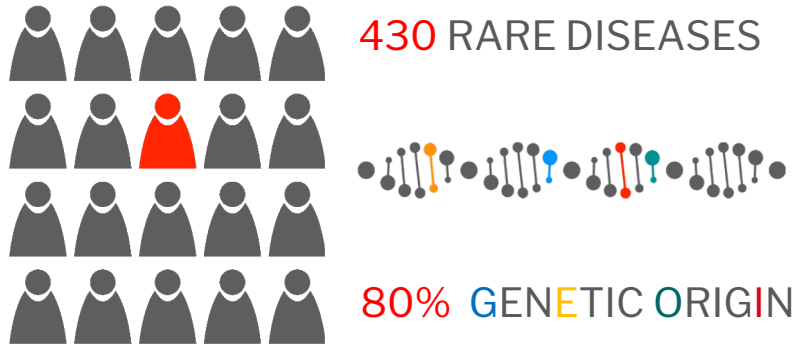
The DataDiVR

Part I: Networks of autoimmunity and autoinflammation

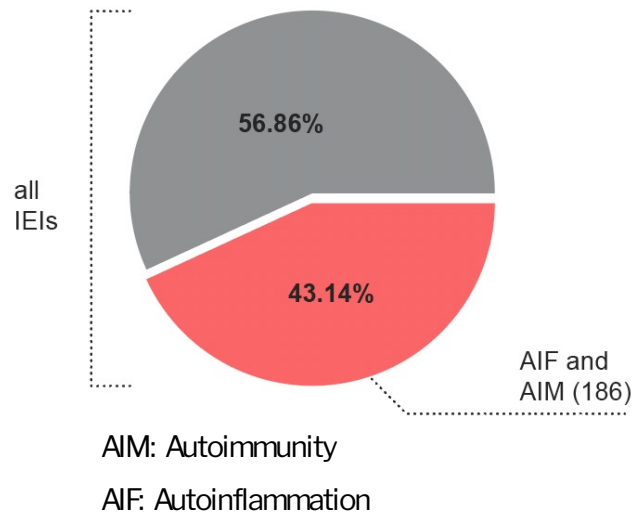


Networks of autoimmunity and
autoinflammation

Inborn errors of immunity (IEI)



- > Collective prevalence ~1/1000–1/5000
- > Severe phenotypes, often from birth
- > 43% have autoimmunity or autoinflammation
- > Strong genotype – phenotype link
- > Insights into immune homeostasis, basis of mechanistic therapies



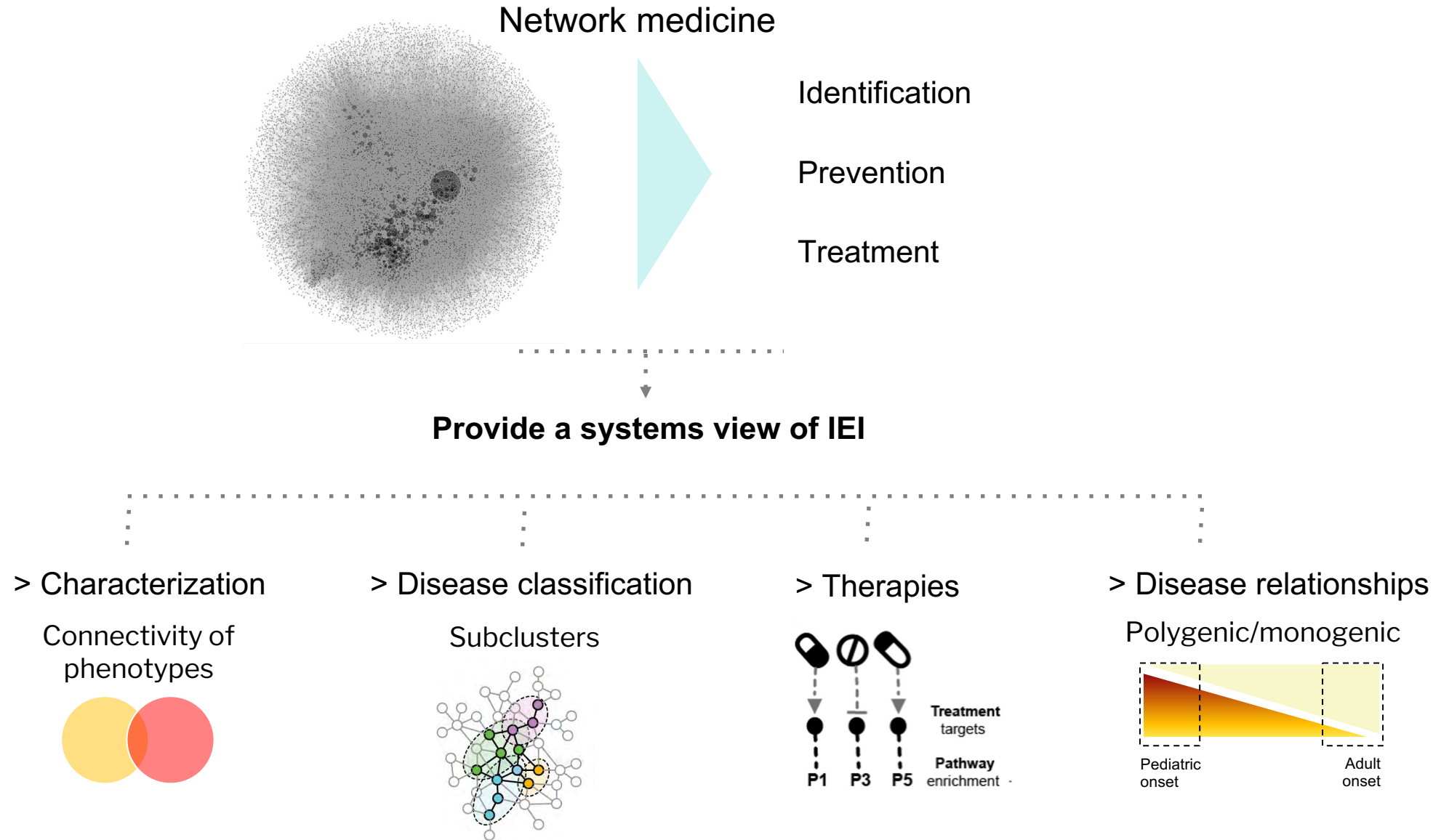
Challenges for IEI

- > Missing heritability
- > Delayed/missing diagnosis
- > Non-specific therapies
- > Poorly characterized

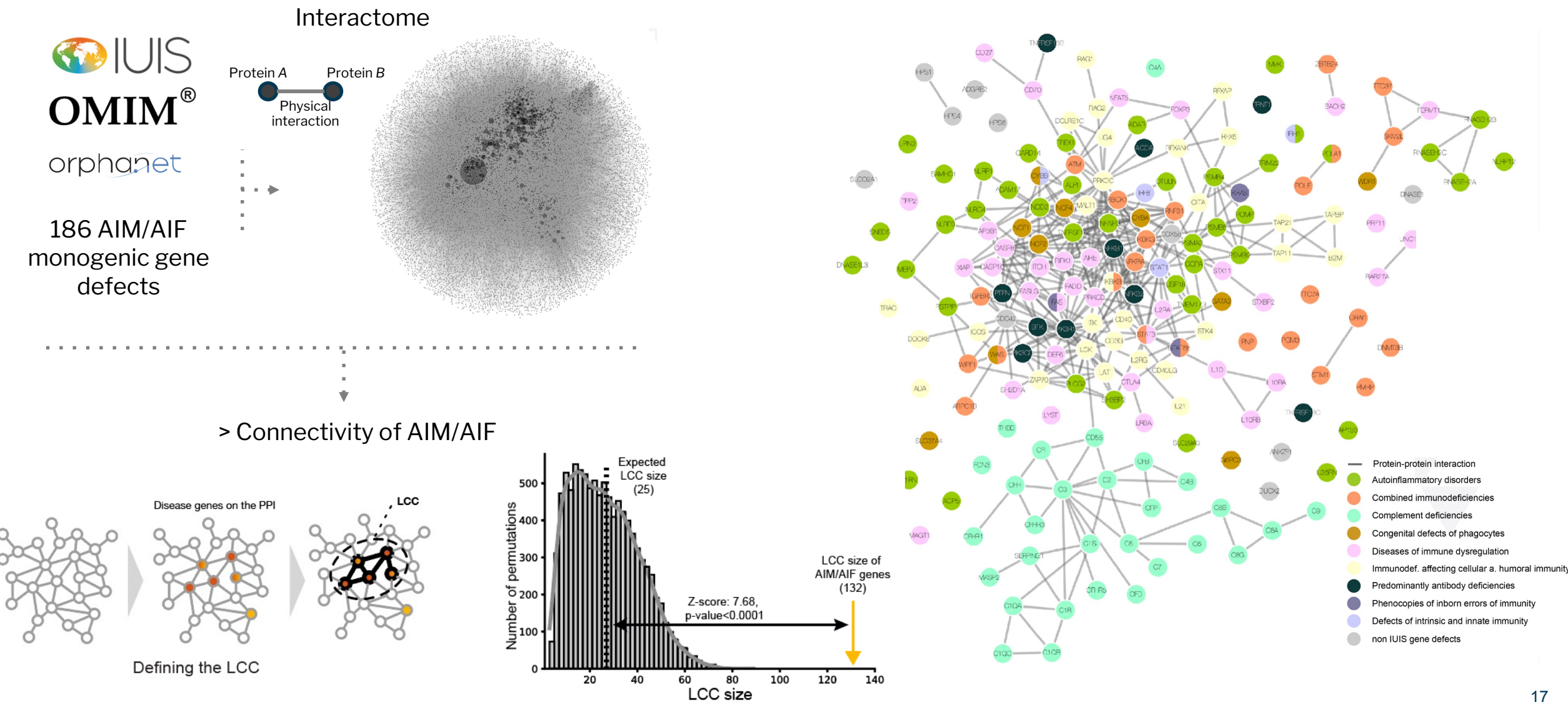
Standardized information

Systematic overview

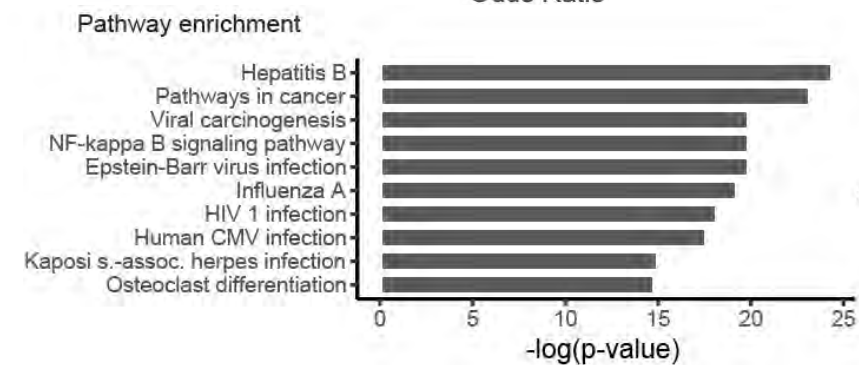
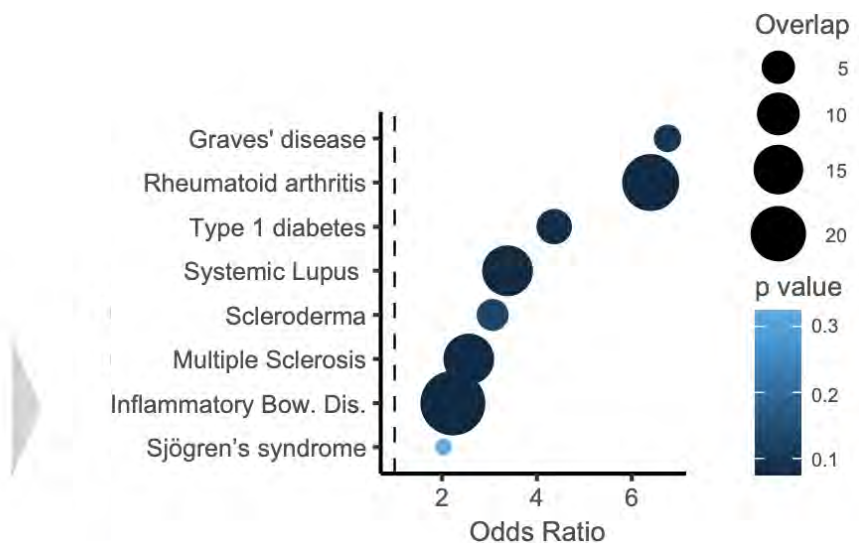
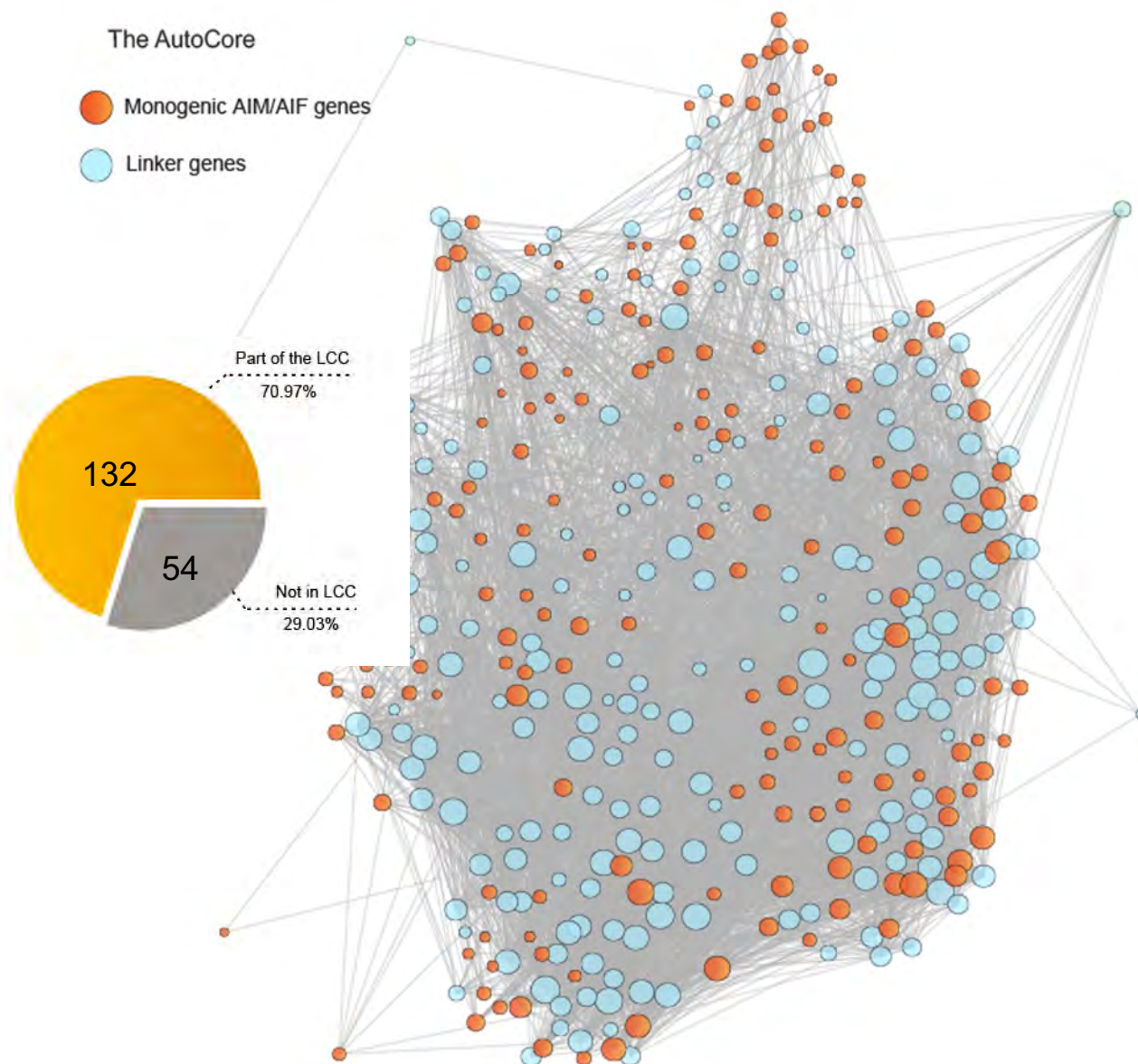
Network medicine for IEI



Monogenic AIM/AIF form a connected cluster

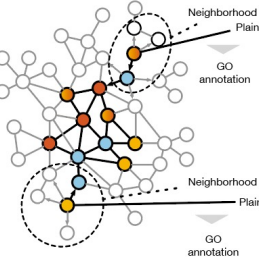


Finding the AutoCore



Subclusters within the AutoCore

AutoCore - GO annotation



Enrichment



Fisher's exact test

P-value < 0.05

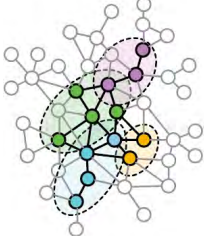
node1 node4

GO 0023273 GO 0023273
GO 0000010 GO 0989993
GO 0535200

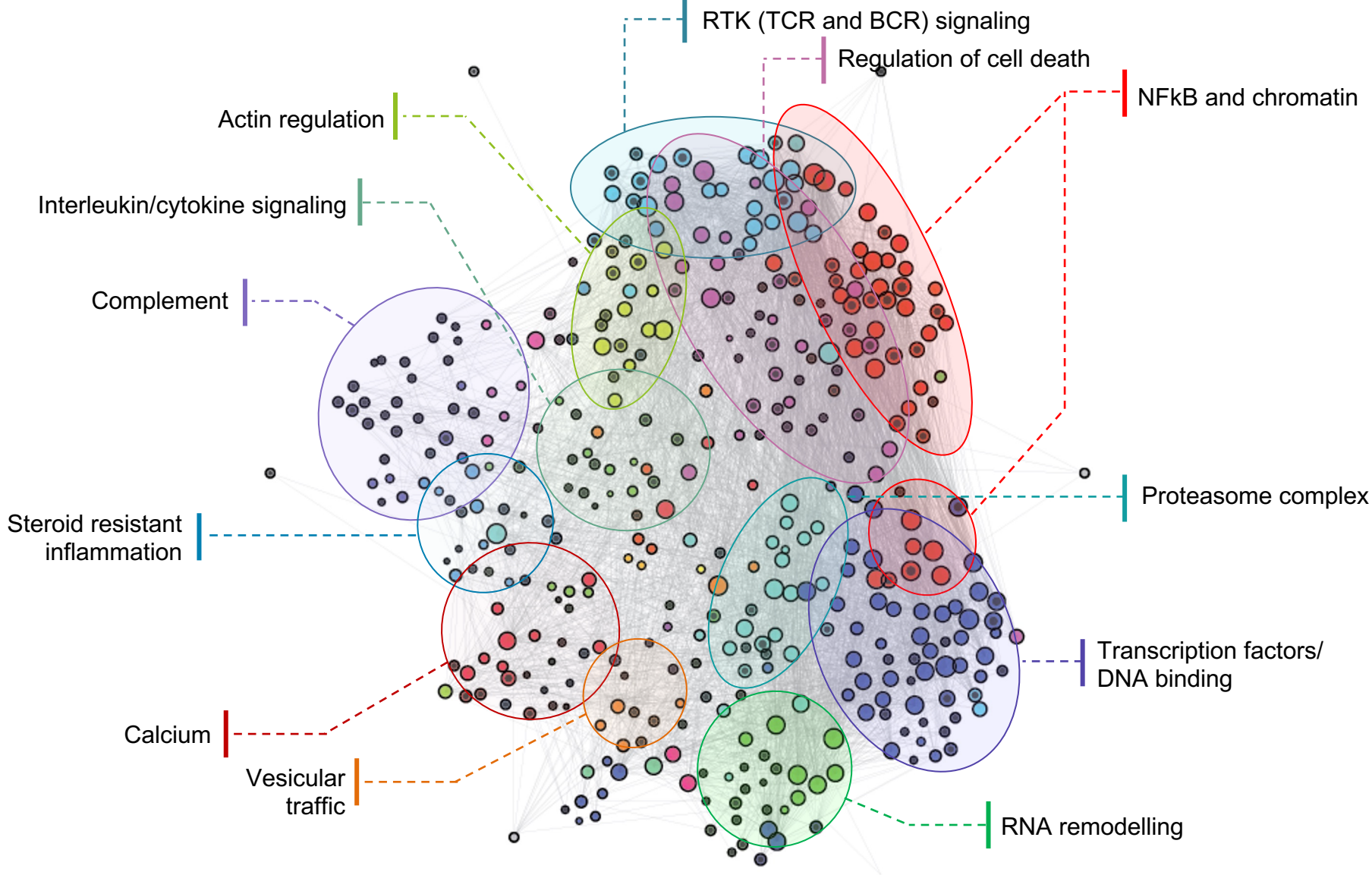
Clustering

node1	node2	node3	node4	node5	node6	
X	X	0	0	X	X	GO 0023273
X	0	X	0	0	0	GO 0000010
X	X	0	0	X	0	GO 0535200
0	X	0	0	X	X	GO 0021148
0	0	0	0	X	X	GO 0989993

Functional atlas

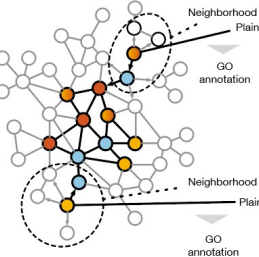


19 disease clusters



Subclusters within the AutoCore

AutoCore - GO annotation



Enrichment



Fisher's exact test

P-value < 0.05

node1 node4

GO 0023273 GO 0023273

GO 0000010 GO 0989993

GO 0535200

19 disease clusters

Clustering



X X 0 0 X X GO 0023273

X 0 X 0 0 0 GO 0000010

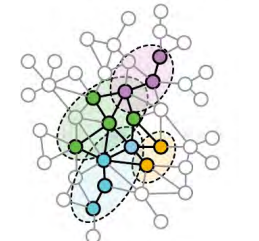
X X 0 0 X 0 GO 0535200

0 X 0 0 X X GO 0021148

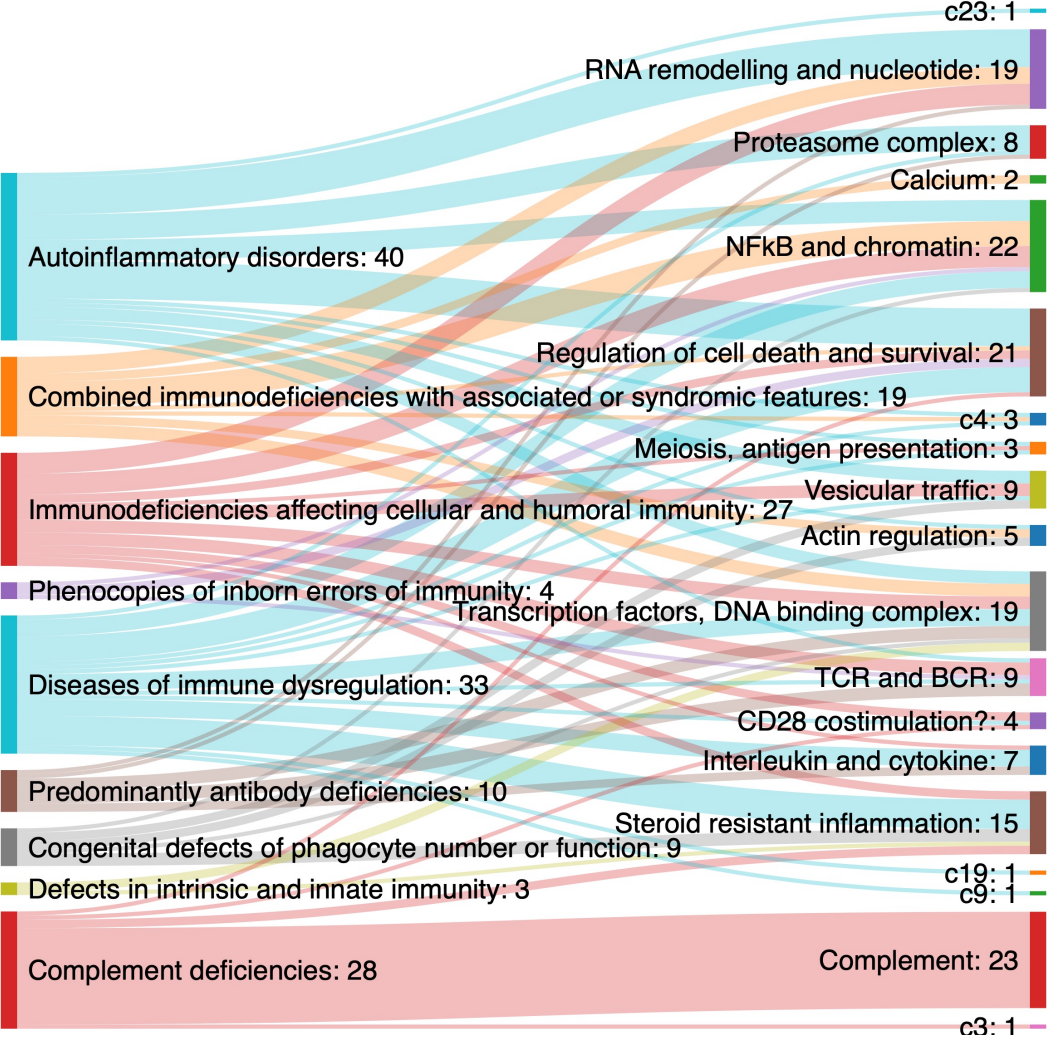
0 0 0 0 X X GO 0989993



Functional atlas



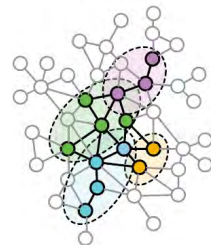
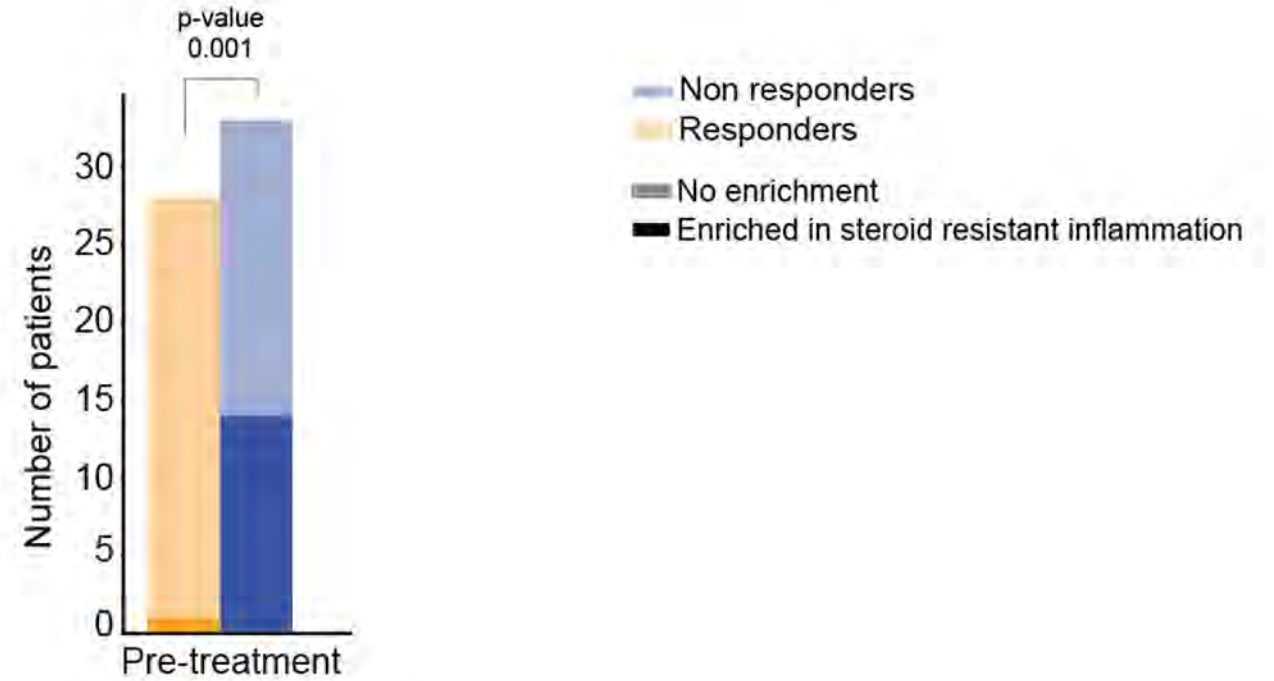
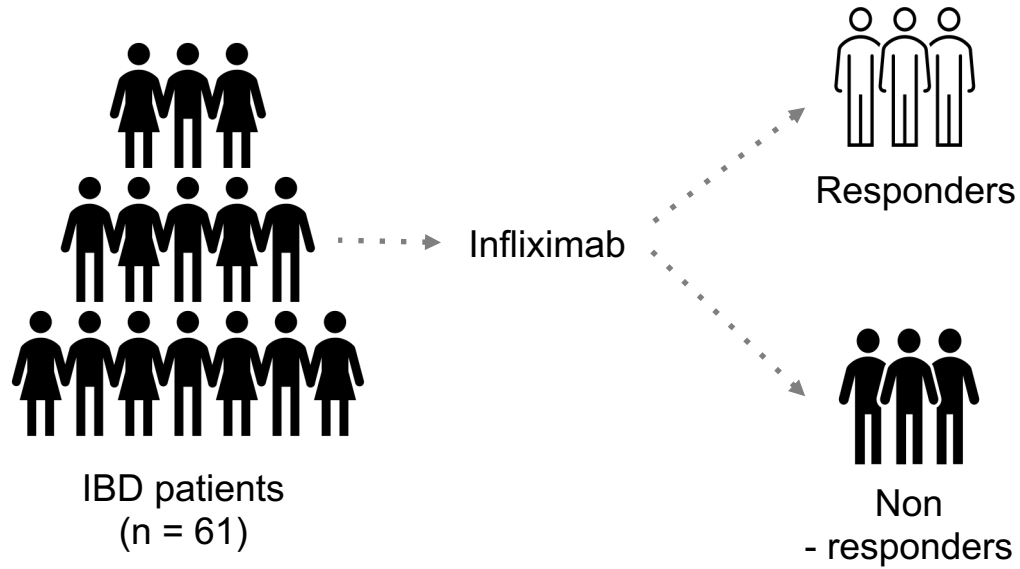
Clinical Clusters (9)



Molecular Clusters (19)

Autocore endotypes differentiate therapy response

> Can the molecular endotypes inform us on response to therapy?



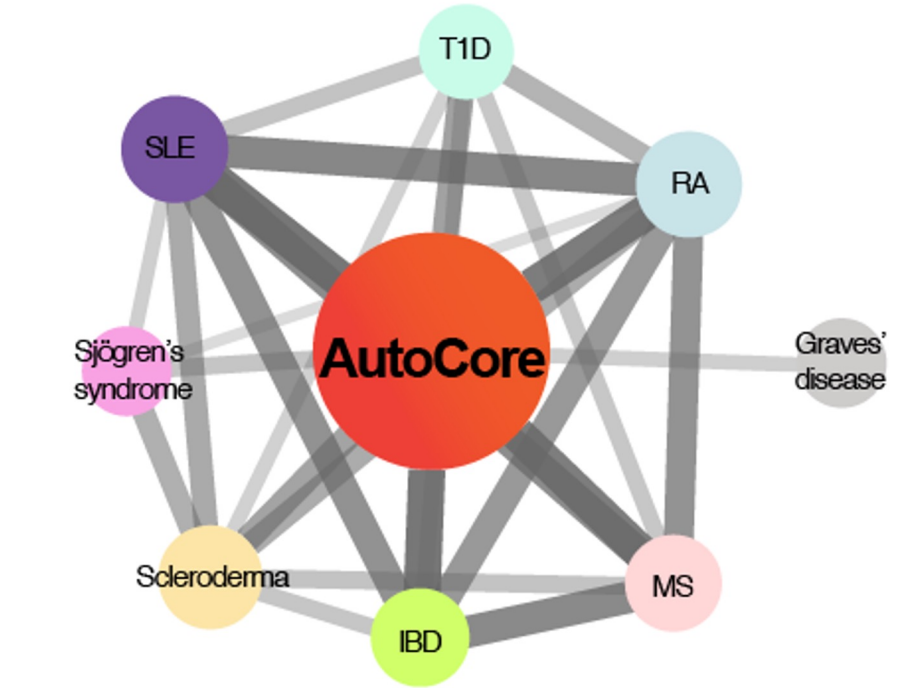
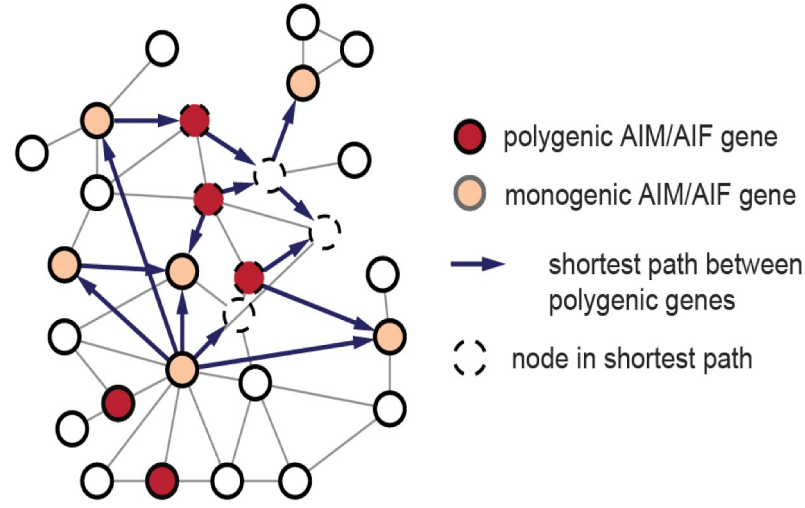
Differential expression

Using the AutoCore to investigate disease relationships

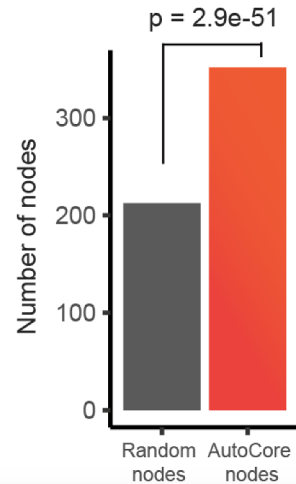
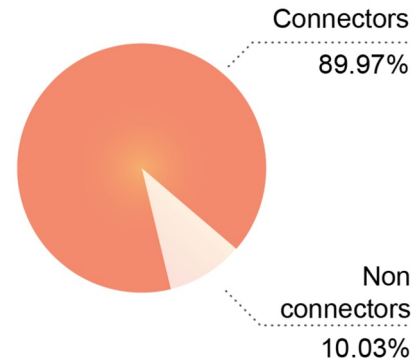
> AutoCore nodes are connectors of polygenic AIM AIF

> Monogenic – polygenic disease network

Graves' disease
 Multiple Sclerosis (MS)
 Inflammatory Bowel Disease (IBD)
 Scleroderma
 Sjögren's Syndrome
 Systemic Lupus Erythematosus (SLE)
 Type 1 Diabetes (T1D)
 Rheumatoid Arthritis (RA)



> All AutoCore nodes



Node size
 > Betweenness centrality



Edge width
 > Mean shortest distance

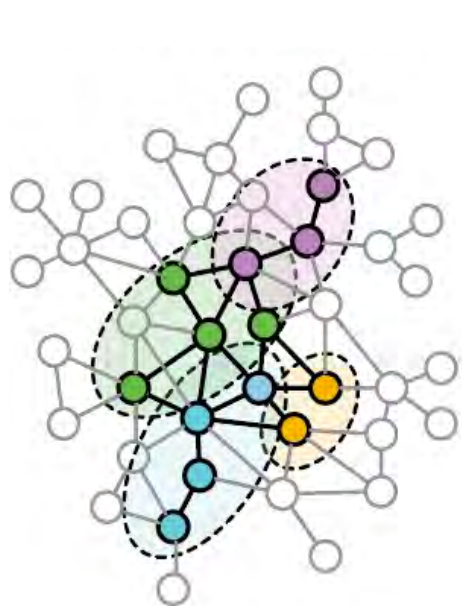


Part II: The DataDiVR

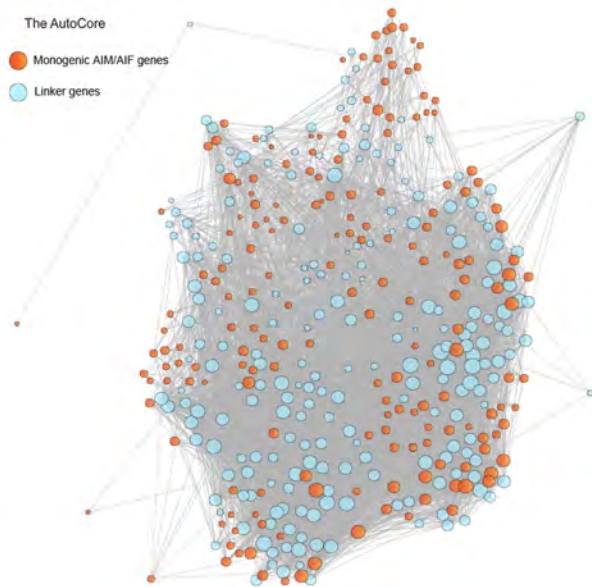


The DataDiVR

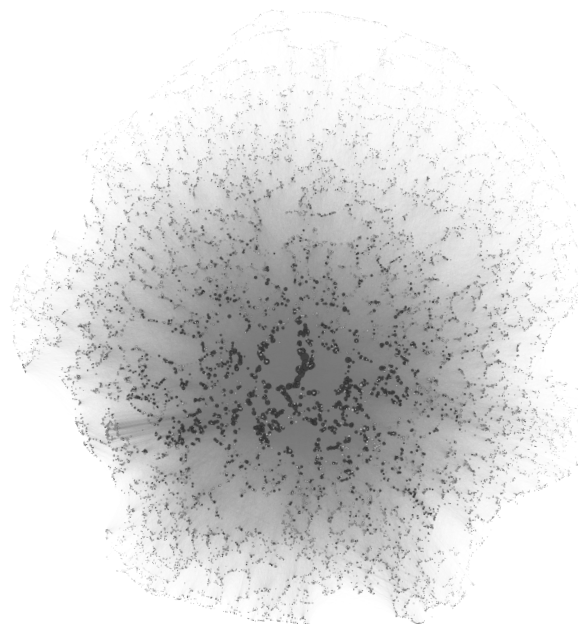
Biological networks in 3D



What you've seen



What I you've seen 2.0



What biological networks look like

```
reading in networks
In [1]: G_full=read_network_csv("../networks/F01_full_n10k.csv")
        G_physical=read_network_csv("../networks/F01_physical_n10k.csv")
        for thenode in G_physical.nodes():
            try:
                sym=ID_symbol[thenode]
            except:
                sym=thenode
            G_physical.nodes[thenode]['sym']=sym
            G_physical.nodes[thenode]['mygroup']='group'

        Gcc = sorted(nx.connected_components(G_physical), key=len, reverse=True)
        GG = G_physical.subgraph(Gcc[5])

> done loading network:
> network contains 19979 nodes and 712242 links
> done loading network:
> network contains 18648 nodes and 304914 links

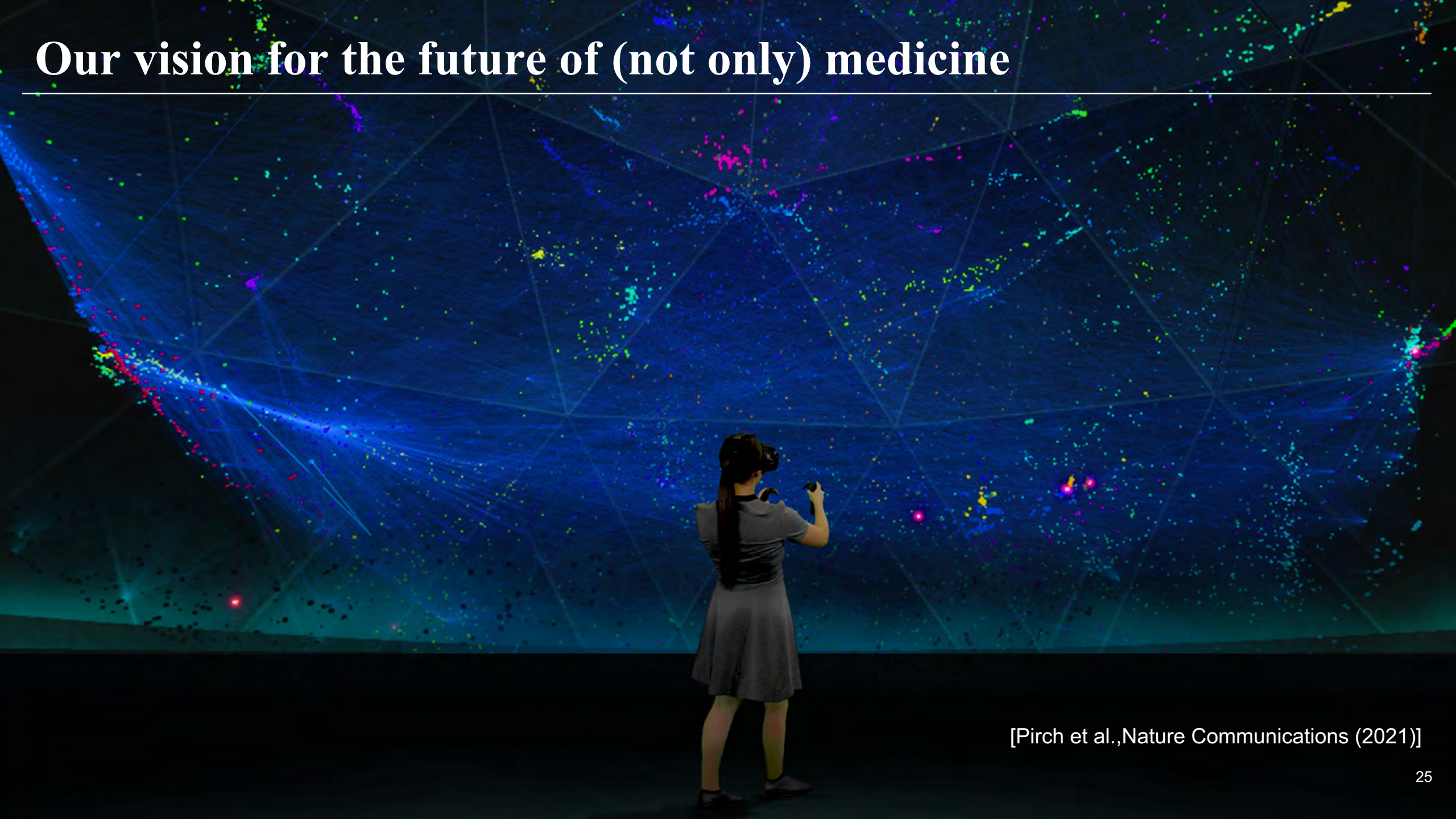
Subnetworks
In [12]: mysubnet=G_physical.subgraph(sorted(mygeneids.values(), []))

        for thenode in mysubnet.nodes():
            group=""
            for mygroup, idlist in mygeneids.items():
                if thenode in idlist:
                    group=mygroup

            try:
                sym=ID_symbol[thenode]
            except:
                sym=thenode
            mysubnet.nodes[thenode]['sym']=sym
            mysubnet.nodes[thenode]['mygroup']=group
```

What working with biological networks looks like

Our vision for the future of (not only) medicine



[Pirch et al., Nature Communications (2021)]

Why Virtual Reality?



Size



Presence



Future

DataDiVR — a Virtual Reality network exploration platform

- + Exploration across scales
- + Exploration across contexts
- + Real-time data analysis
- + Modular and open design
- + Runs on standard hardware or in the cloud
- + Proof-of-concept rare disease variant prioritization



[Pirch et al., Nature Communications (2021)]





Team Jörg Menche

Anna Hakobyan
Chloé Bucheron
Celine Sin
Christiane V. R. Hütter
Daniel Malzl
Felix Müller
Joel Hancock
Julia Pazmandi
Karl Riepl
Loan Vulliard
Martin Chietini
Salvo D Lombardo
Sebastian Pirch
Till Oblau
Varun Sharma

Contact

✉ julia@menchelab.com

🌐 www.menchelab.com

🐦 [@menchelab.com](https://twitter.com/menchelab.com)