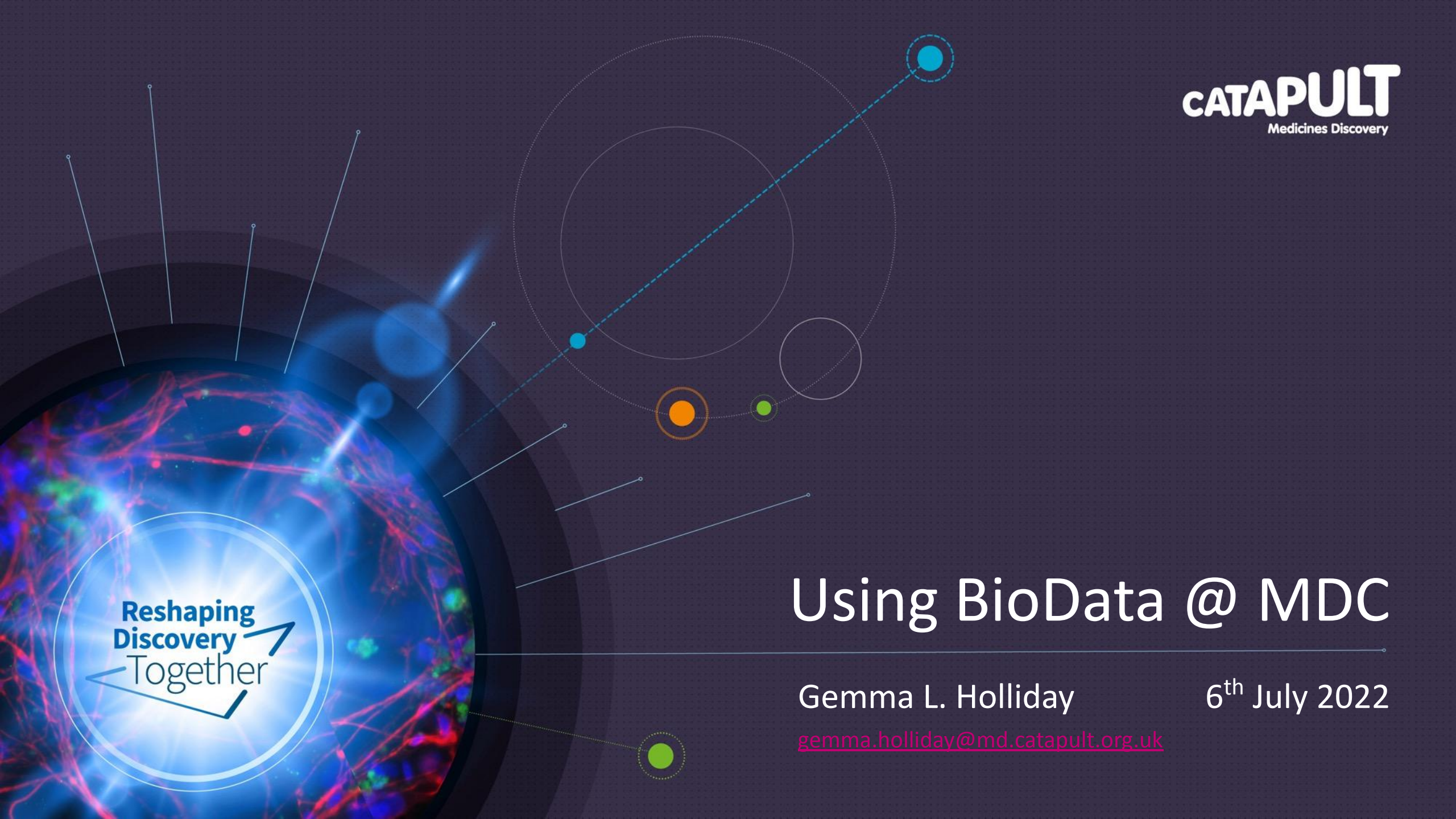


CATAPULT
Medicines Discovery

Using BioData @ MDC

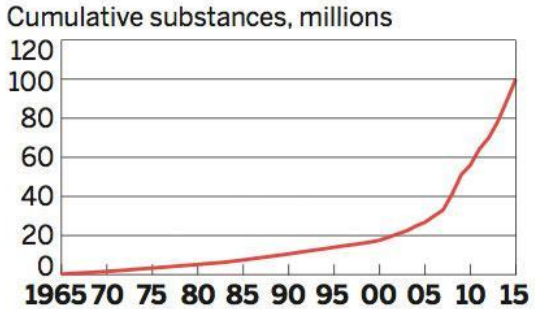
Gemma L. Holliday

gemma.holliday@md.catapult.org.uk

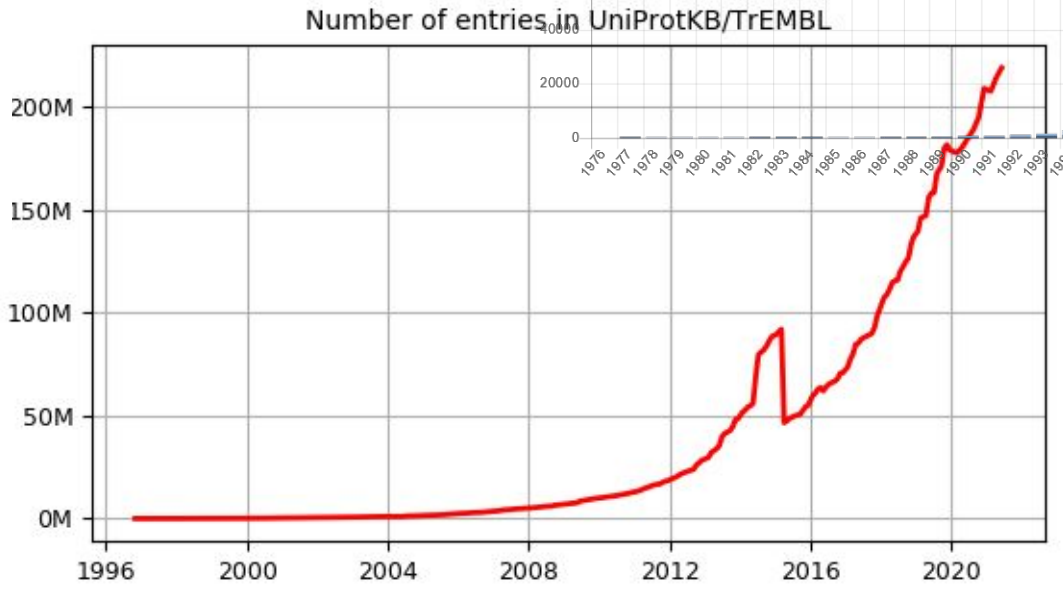
6th July 2022

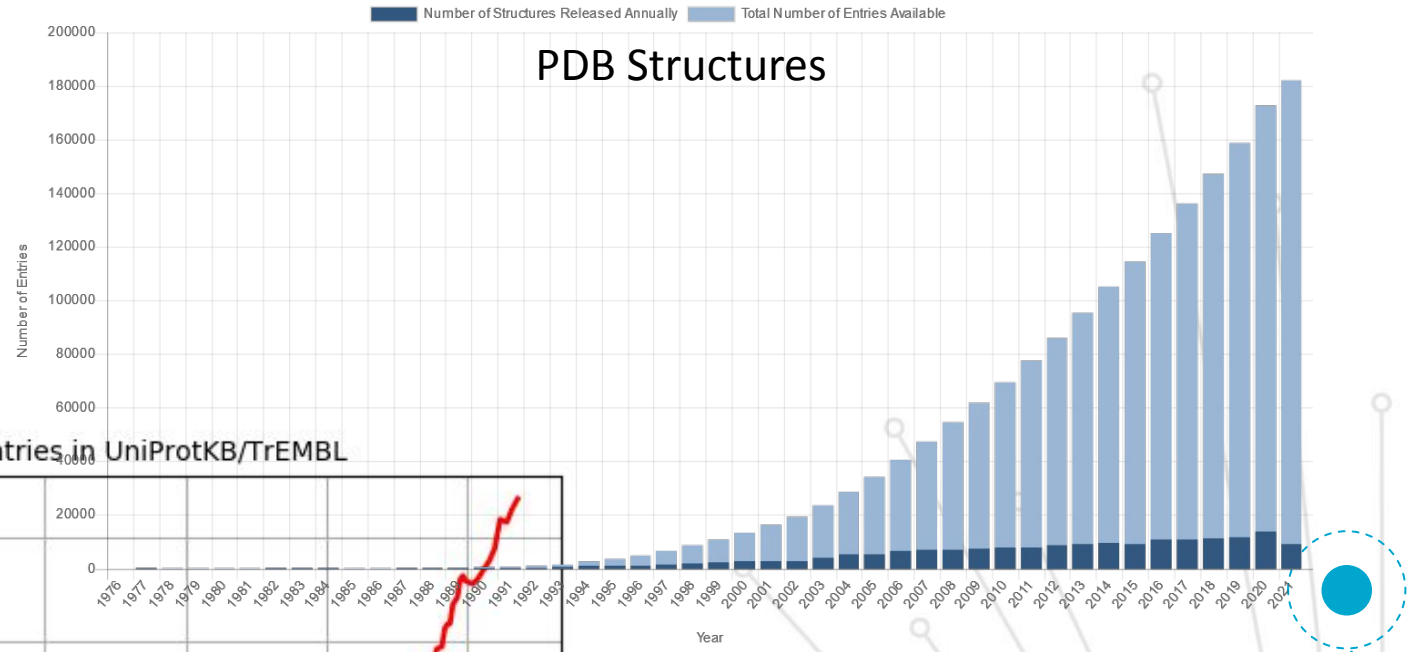
The Data Universe is Expanding

Chemical Abstract Service

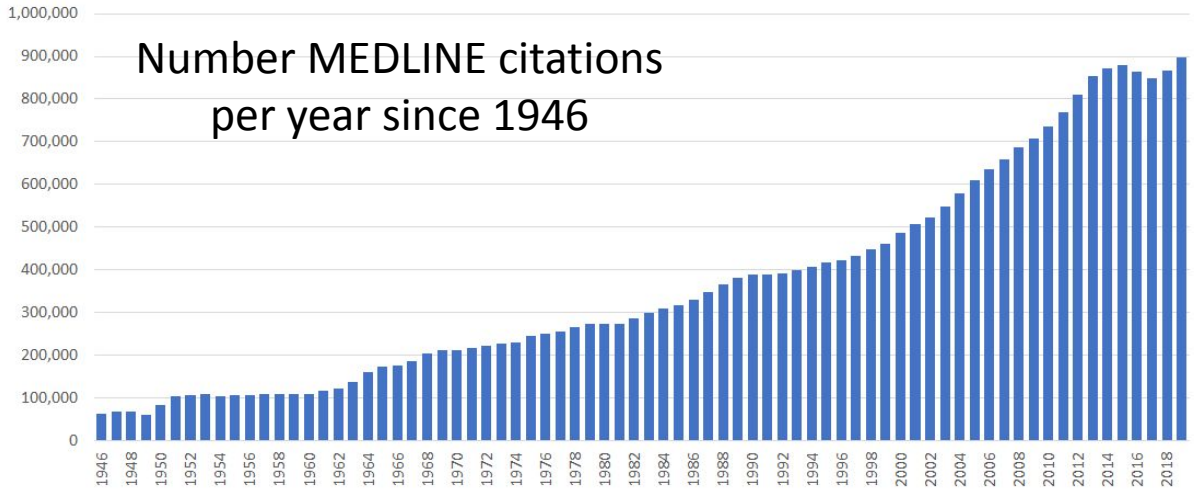


PubChem contains 110,666,946 compounds as of Nov 2021



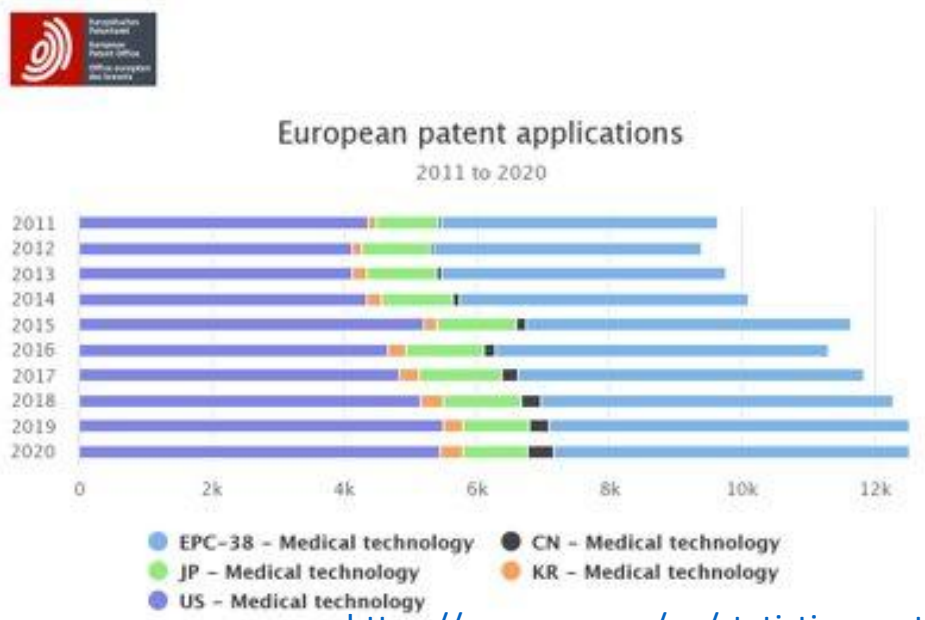


The Data Universe is Expanding



https://dblp.org/statistics/publicationsperyear.html

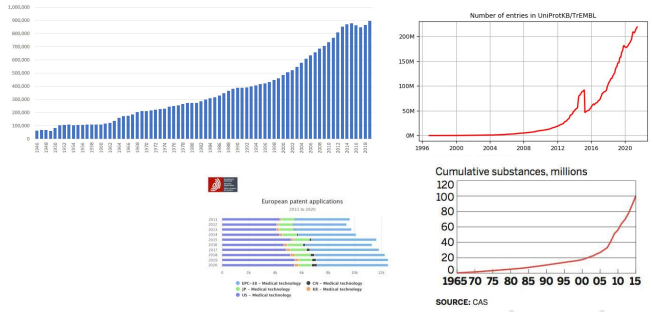
Medline in 2019 had > 27 Million Citations



https://new.epo.org/en/statistics-centre

The Data Universe is Expanding

- The Data Universe is expanding faster than we can keep up

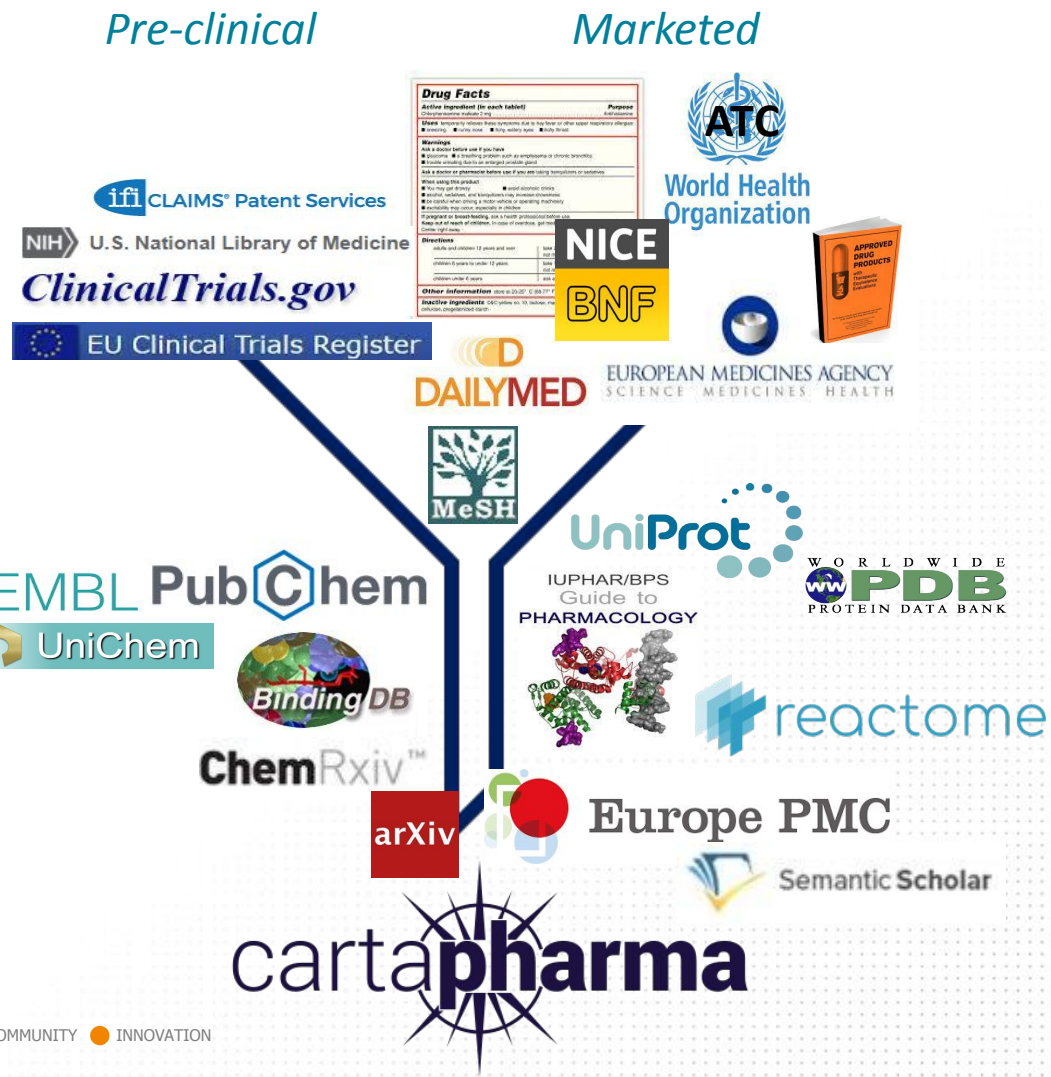


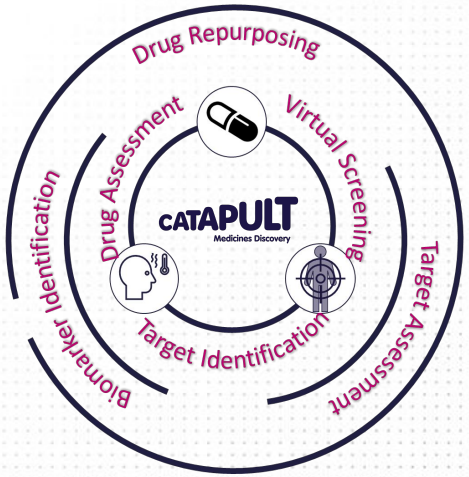
- There is still a lot we don't know!
- In July 2022, there were **231,921,735** proposed proteins:
  - **1,730,144** have evidence at the protein or transcript level (0.7 %)
  - **77,239,270** are inferred via homology (33.3 %)
  - Roughly 2/3rd are "predicted"
  - **567,483** (0.2 %) of these are in the human curated section of UniProtKB (**29.2 %** have evidence at the protein or transcript level)
- Even well studied superfamilies of proteins have vast swathes of members that are simply unknowns
  - Of the ~ 600 kinase of the human kinome, only 140 have more than 100 citations, ~500 fewer than two
    - Based on a EuroPMC search of human kinome "kinase gene_name" data accessed 04/07/2022

We need more than ***just*** data

- Knowledge comes from the understanding of data and how we can use it.
- Expert curation in structured resources (e.g., UniProtKB) is critical
  - But costly and hard to maintain
- Data lakes and data silos (often held in house or behind paywalls)
- Literature and other free text
- Methods, such as NLP, and NER
- Machine learning, and Artificial Intelligence

MDC’s Data Lake





Mining Unstructured Data for Knowledge

Extract entities of interest

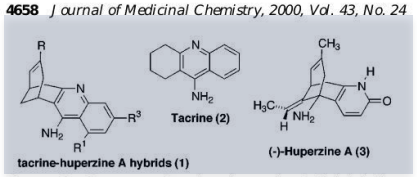
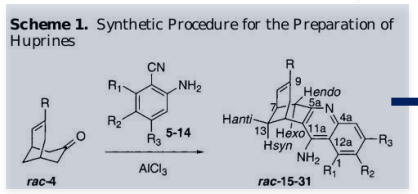


Figure 1. Structure of tacrine–huperzine A hybrids (huprines) and their starting models.

perzine A and 25-fold more potent than tacrine.[13] Also, some compounds designed by dimerization of the same fragment, huperzine A, proved to be about 2-fold more potent than (−)-huperzine A and 4-fold more potent than tacrine.[14] Some galanthamine-based heterodimers were also up to 5-fold more potent than tacrine and up to 36-fold more potent than galanthamine.[15] Another bis-interacting ligand, designed by combining fragments of the structures of huperzine A and donepezil, has also been recently synthesized, although this compound showed no effective AChE inhibitory activity.[16]

Recently we have reported the synthesis, in vitro pharmacology, and molecular modeling of a series of tacrine–huperzine A hybrids (huprines) of general structure **1** (Figure 1) as AChE inhibitors of potential interest for the treatment of AD.[17−19] These compounds were originally designed in an empirical way by combination of the pharmacophores of huperzine A (carbobicyclic substructure) and tacrine (4-aminoquinoline substructure) to improve their binding to the active site of AChE. The structure of these compounds do not seem adequate to simultaneously bind to both the active sites and the peripheral sites of AChE. Several of these compounds exhibited higher AChE inhibitory activity than tacrine (**2**) and (−)-huperzine A (**3**) (Figure 1), particularly when a methyl (*rac*-**15**) or ethyl (*rac*-**21**) group was attached to position 9. Moreover, the introduction of a fluorine substituent at position 3 (*rac*-**18**) was also found to be advantageous, leading to a compound 15 times more active than tacrine in inhibiting AChE from bovine erythrocytes.[17] Likewise, the AChE inhibitory activity of the levorotatory enantiomers was roughly twice that of the racemic mixtures, while the dextrorotatory enantiomers were much less active. Molecular modeling of the interaction of these compounds with *Torpedo californica* AChE (*Tc*AChE) suggested that they behave as true tacrine–huperzine A hybrids, since the 4-aminoquinoline and bicyclo[3.3.1]-nonadiene subunits roughly occupy the same positions of the corresponding moieties in tacrine and (−)-huperzine A, respectively, as determined from their crystallographic complexes with AChE.[17,18] Later, replacement of fluorine by chlorine at position 3 (*rac*-**30**) was found to improve the inhibitory activity, leading to an inhibition constant (K_i) for human AChE around 30 pM, which means an affinity around 1200-fold higher than



Scheme 1. Synthetic Procedure for the Preparation of Huprines

	R
rac-**4a**	CH_3
rac-**4b**	CH_2CH_3

	R_1	R_2	R_3
5	CH_3	H	H
6	H	H	CH_3
7	CH_3	H	CH_3
8	F	H	H
9	H	H	F
10	F	H	F
11	Cl	H	H
12	H	Cl	H
13	H	H	Cl
14	Cl	H	Cl

	R	R_1	R_2	R_3
15	CH_3	H	H	H
16	CH_3	H	H	CH_3
17	CH_3	F	H	H
18	CH_3	H	H	F
19	CH_3	F	H	F
20	CH_3	H	H	Cl
21	CH_2CH_3	H	H	H
22	CH_2CH_3	CH_3	H	H
23	CH_2CH_3	H	H	CH_3
24	CH_2CH_3	CH_3	H	CH_3
25	CH_2CH_3	F	H	H
26	CH_2CH_3	H	H	F
27	CH_2CH_3	F	H	F
28	CH_2CH_3	Cl	H	H
29	CH_2CH_3	H	Cl	H
30	CH_2CH_3	H	H	Cl
31	CH_2CH_3	Cl	H	Cl

positions (1, 2, 3, or 1,3) of the benzene ring. The pharmacological analysis includes: (a) inhibitory activity of bovine and human AChE and human butyrylcholinesterase (BChE), (b) neuromuscular studies focused on the ability to revert the neuromuscular blockade induced by *d*-tubocurarine, (c) time dependence and reversibility of the AChE inhibitory activity, and (d) ex vivo AChE inhibitory activity studies. Finally, molecular modeling studies have been performed to explain the differences in inhibitory activity of the more active compounds.

Results and Discussion

Chemistry. The synthesis of the new compounds (*rac*-**19**, *rac*-**24**–*rac*-**29**, and *rac*-**31**) was carried out by Friedländer reaction of the known enones *rac*-**4a** and *rac*-**4b**[21] and the corresponding aminobenzonitrile **7**–**12** or **14**, under aluminum trichloride catalysis, usually in 1,2-dichloroethane as solvent under reflux. Enones *rac*-**4a** and *rac*-**4b** were easily prepared from the commercially available bicyclo[3.3.1]nonane-3,7-dione by reaction with the appropriate organomagnesium, organolithium, or organocerium reagent to give a 3-alkyl-2-oxa-1-adamantanol, which was then mesylated and submitted to a silica gel promoted fragmentation reaction. Aminobenzonitriles **7**,[22] **9**,[23] **10**,[24] **11**,[25] **12**,[26] and **14**[27] were prepared by the described procedures, while **8** is a commercial compound. Not unexpectedly on steric grounds, the yield of these reactions was low in the cases where the product contained a chlorine (*rac*-**28** and *rac*-**31**) or methyl (*rac*-**24**) substituent at position 1, despite

Extract images and index the labels

Image

C1C(=O)CCCC1

Chemical Representation

Embedded tables to CSV

A potential role for PKN1 in cytoskeletal and mitochondrial disruption in ALS.

Related Entities

Drug Repurposing in Healthy Ageing

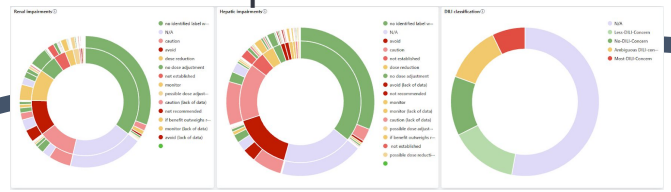
- The **Challenge**: Successful drug repositioning in the elderly is challenging due to a multitude of safety, efficacy and tolerability factors
- **Aim**: To utilise informatics approaches to identify subset of approved drugs that are appropriate for realistic translation to patient population for aging interventions
- Visit **www.DR-EAM.ai** -- **D**rug **R**epurposing for **E**lderly **A**nd **M**ultimorbid

- Supported by cartapharma
- Data indicative of the age-related appropriateness for approved drugs world-wide (including NHS data)
- Data mining, Natural Language Processing, and Machine Learning approaches used
- Interactive web interface developed to allow selection and prioritisation of drugs based on user preference for repositioning in an elderly population.

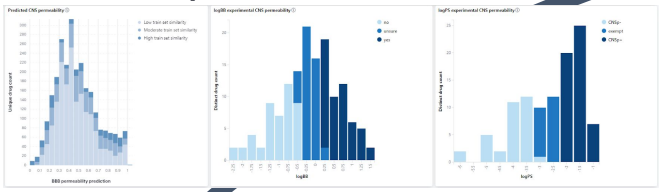
CATAPULT Medicines Discovery

UK Spine Knowledge Exchange

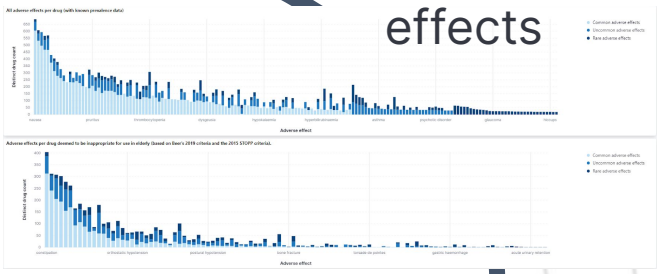
Renal & hepatic impairments



CNS penetration



Adverse effects

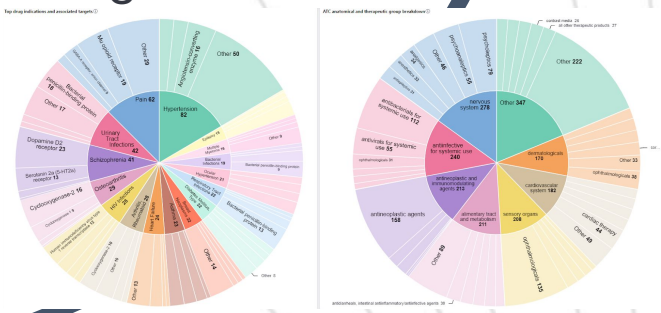


Making all the information available to make the best decisions in real time

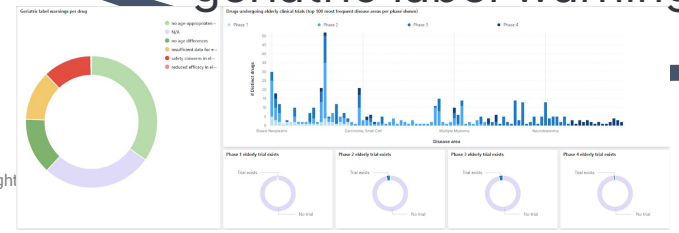
Drug-drug interactions

Thetis

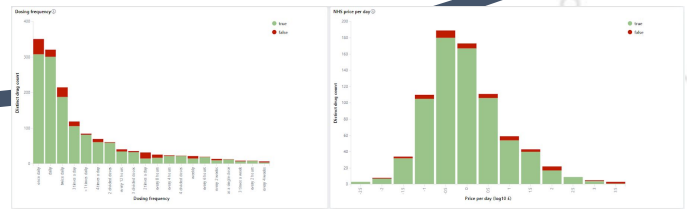
Target information

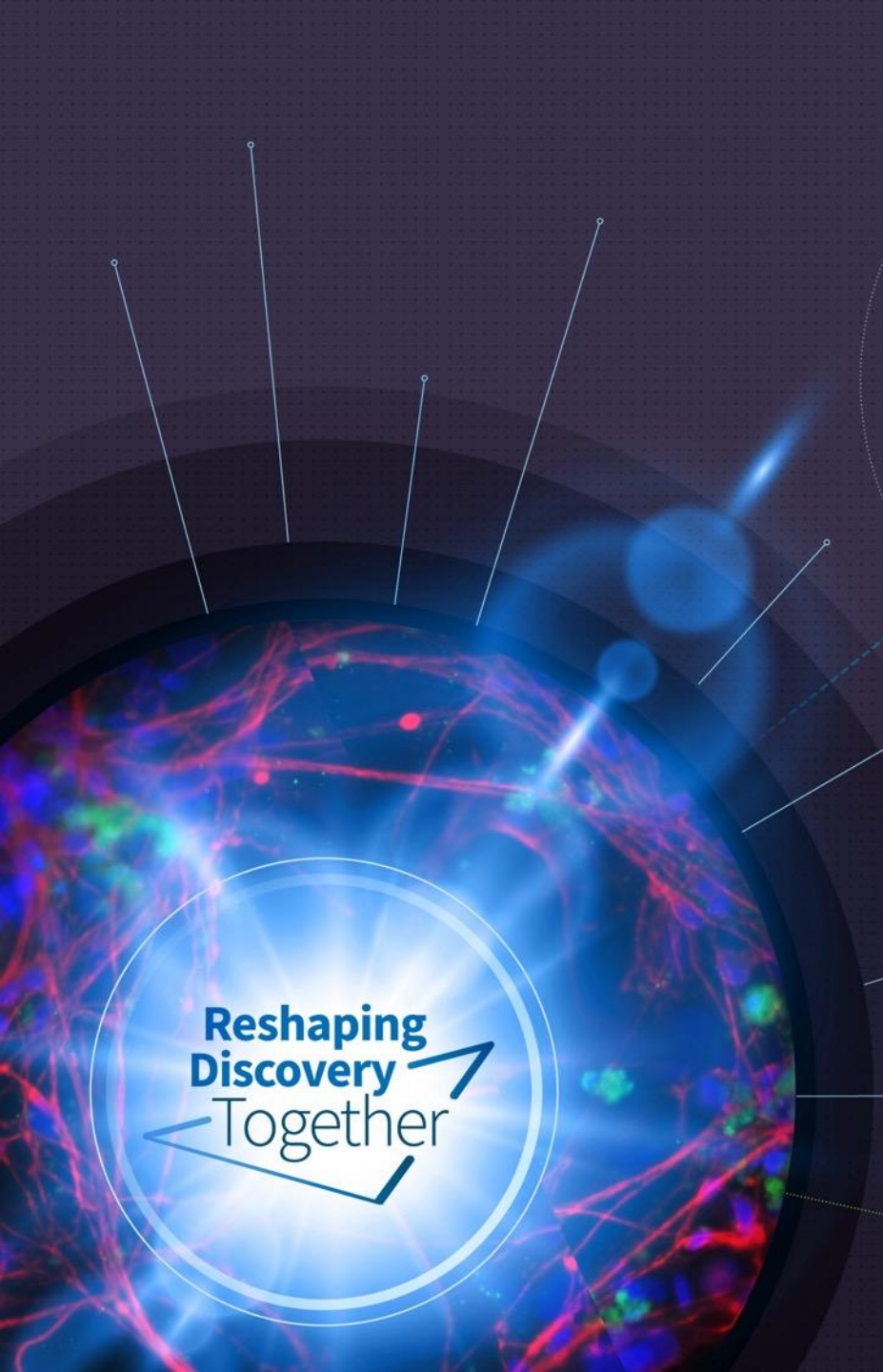


Elderly clinical trials & geriatric label warnings



Daily dosing information





UK Biobank and Genomics England — MDC Experience: Challenges and opportunities

Andrey Gagunashvili

6th July 2022

UK Biobank vs Genomics England

UK Biobank

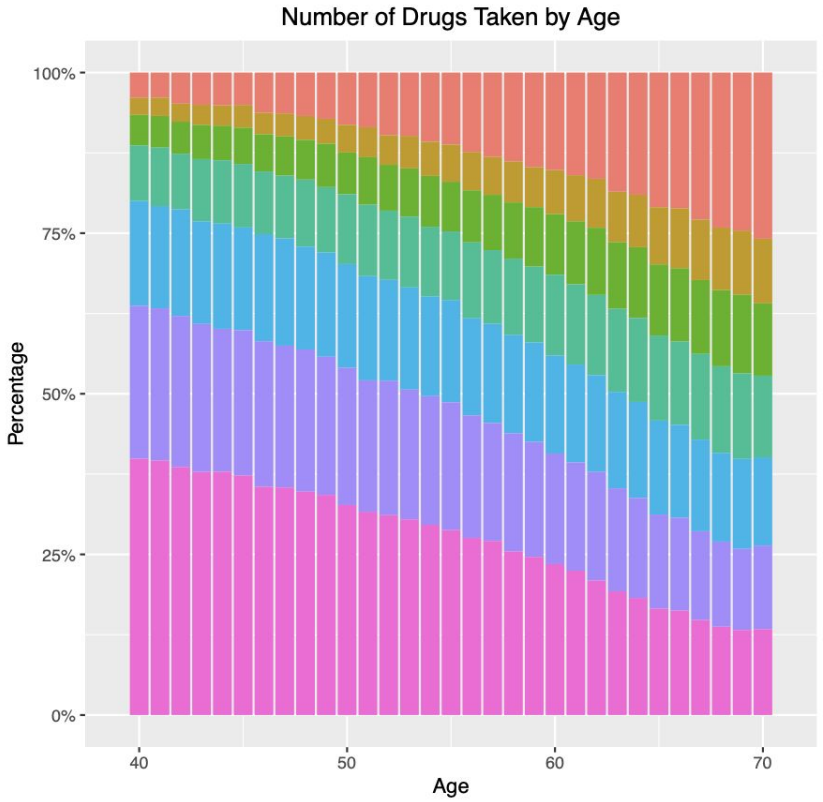
- Mostly "healthy" adults
- Whole exome and whole genome sequencing
- May contain rare disease-causing variants in late-onset disease genes
- Can be used as a "healthy", control cohort, similar to gnomAD

Genomics England

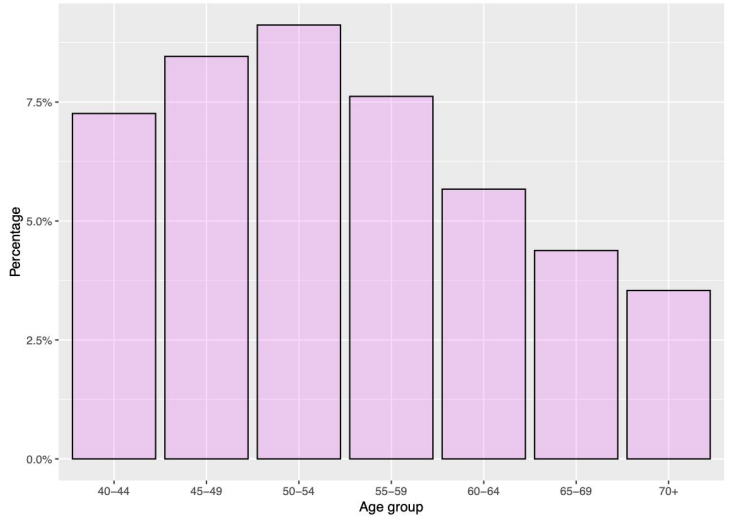
- Rare diseases (early and late onset) and cancer
- Whole genome sequencing
- Enriched with "rarer" variants and genes that were missed during prior genetic testing/screening for "common" disease genes

UK Biobank

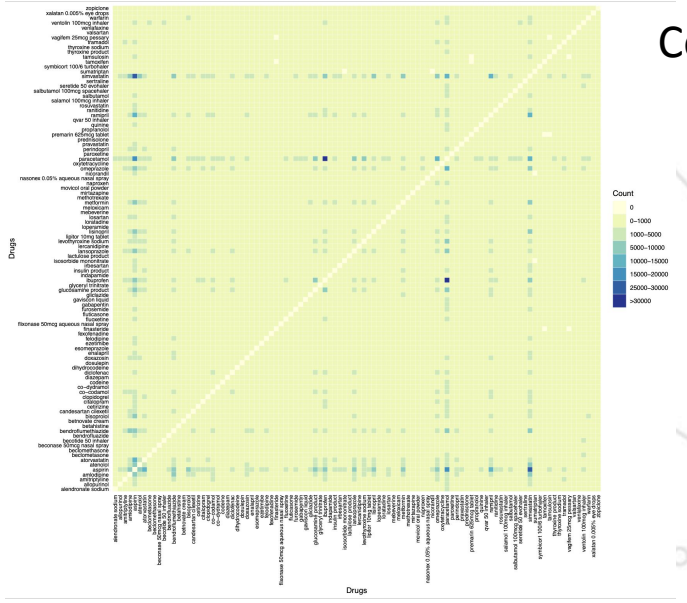
- ***"Drug reporposing for elderly population"***
- Survey of drugs currently used in ageing patients
- Polypharmacy and drug co-prescription



The older we get the more drugs we take



Estrogen drugs

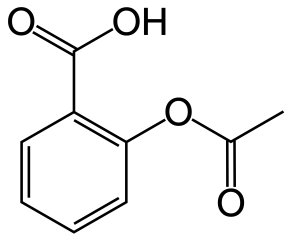


Co-prescribed drugs

UK Biobank

Challenges:

- Very many information categories/fields
- Can be sketchy, not complete
- Not unified naming of some of the administered drugs, e.g. based on the active compound:



acetylsalicylic acid
aspirin
aspirin 75 mg tablet
aspirin+codeine
aspirin+codeine 300mg/8mg tablet
alka-seltzer tablet

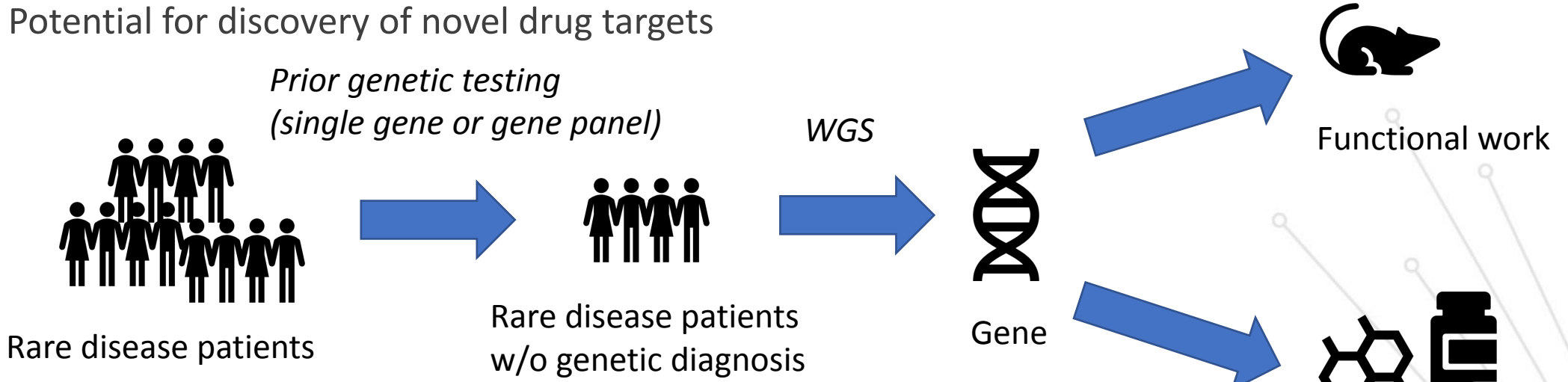
Opportunities:

- MDC has an expertise to standardise drug names based on the active ingredient, e.g. INN name:

acetylsalicylic acid → ***aspirin***
aspirin → ***aspirin***
aspirin 75 mg tablet → ***aspirin***
aspirin+codeine → ***aspirin***
aspirin+codeine 300mg/8mg tablet → ***aspirin***
alka-seltzer tablet → ***aspirin***

Genomics England

- Rare disease cohorts
- "Enriched" for "rarer" disease genes
- Potential for discovery of novel drug targets



Genomics England

Challenges:

- Commercial use requires a commercial license
- Closed research environment
- Not possible to install tools and datasets on your own
- Dependance on the GEL help desk
- Not always up-to-date, e.g. ClinVar
- Patient's characteristics are subjective to the recruiting clinician, e.g:
  - Diagnosis
  - Phenotype description (e.g. HPO terms)

Opportunities:

- Virtual machine environment allowing researchers to install their own tools and datasets

Conclusions

- Data are going to keep growing
  - More genetic & phenotypic data
  - More sequences and protein data
  - More chemical and activity data
- Data Knowledge remains challenging
- Increasingly complex
- We need to:
  - Be able to translate between resources
  - Understand underlying bias in the data
  - Find ways to keep up-to-date
- Integration of resources is critical

Acknowledgements

- Current Team Members
  - Rafael Jimenez
  - Ian Dunlop
  - Matthew Hodgskiss
  - Mouhamad Aboshokor
  - Dan James
  - Roman Ma
  - Kepa Burusco-Goni
  - Samrina Rehman
- Previous Team Members
  - John P. Overington
  - Andrew Pannifer
  - Charles Burry
  - Anna Pallo
- Our Funders