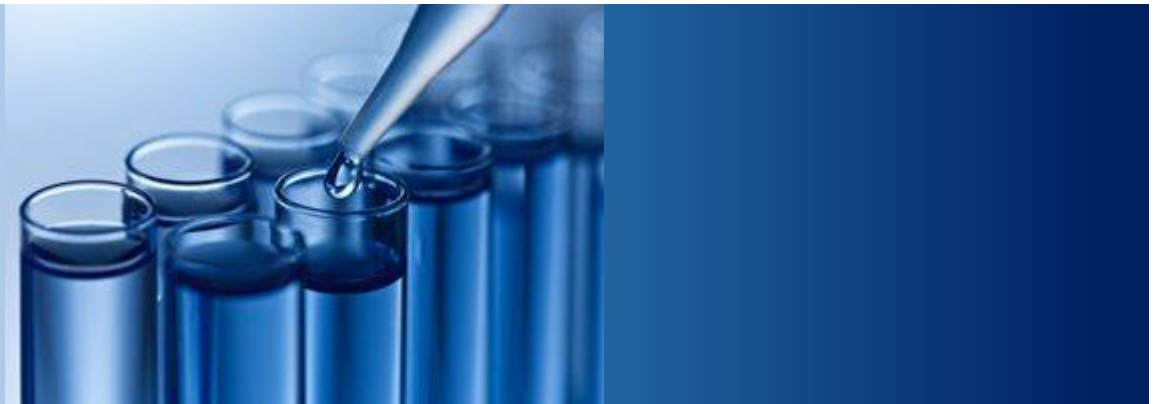


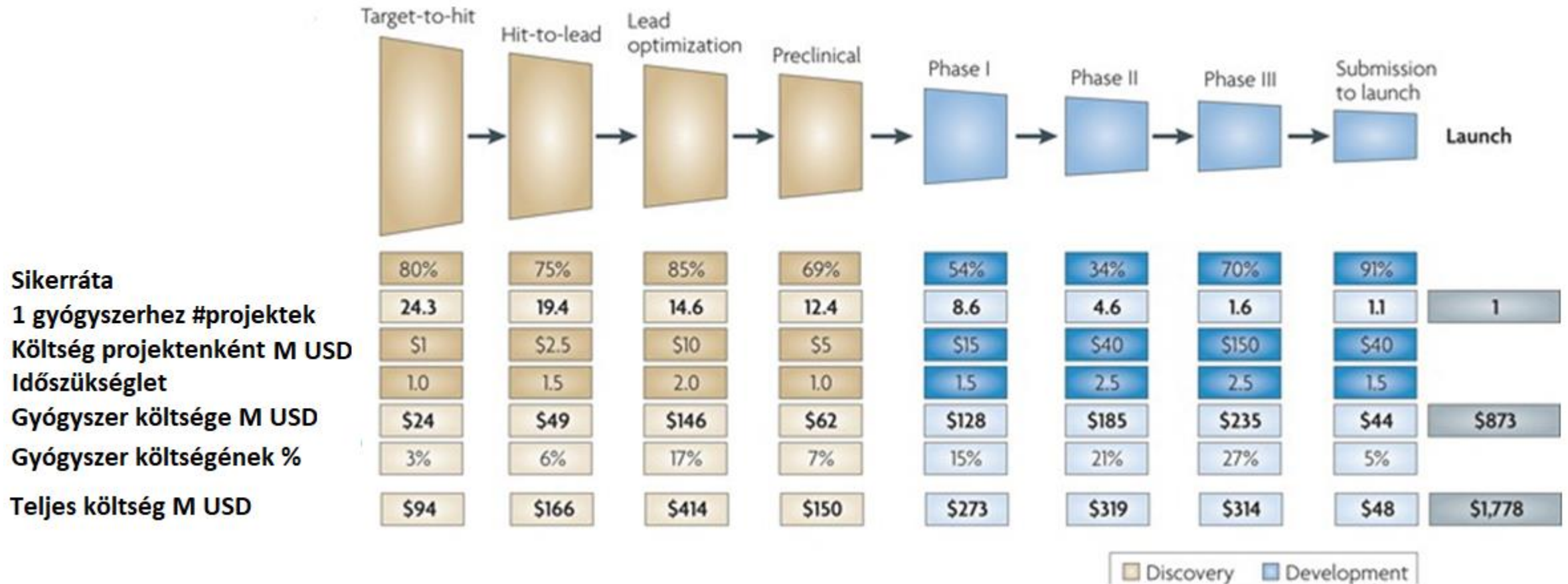
# Mesterséges intelligencia a gyógyszerkutatásban



Keserű György Miklós és Ferenczy György

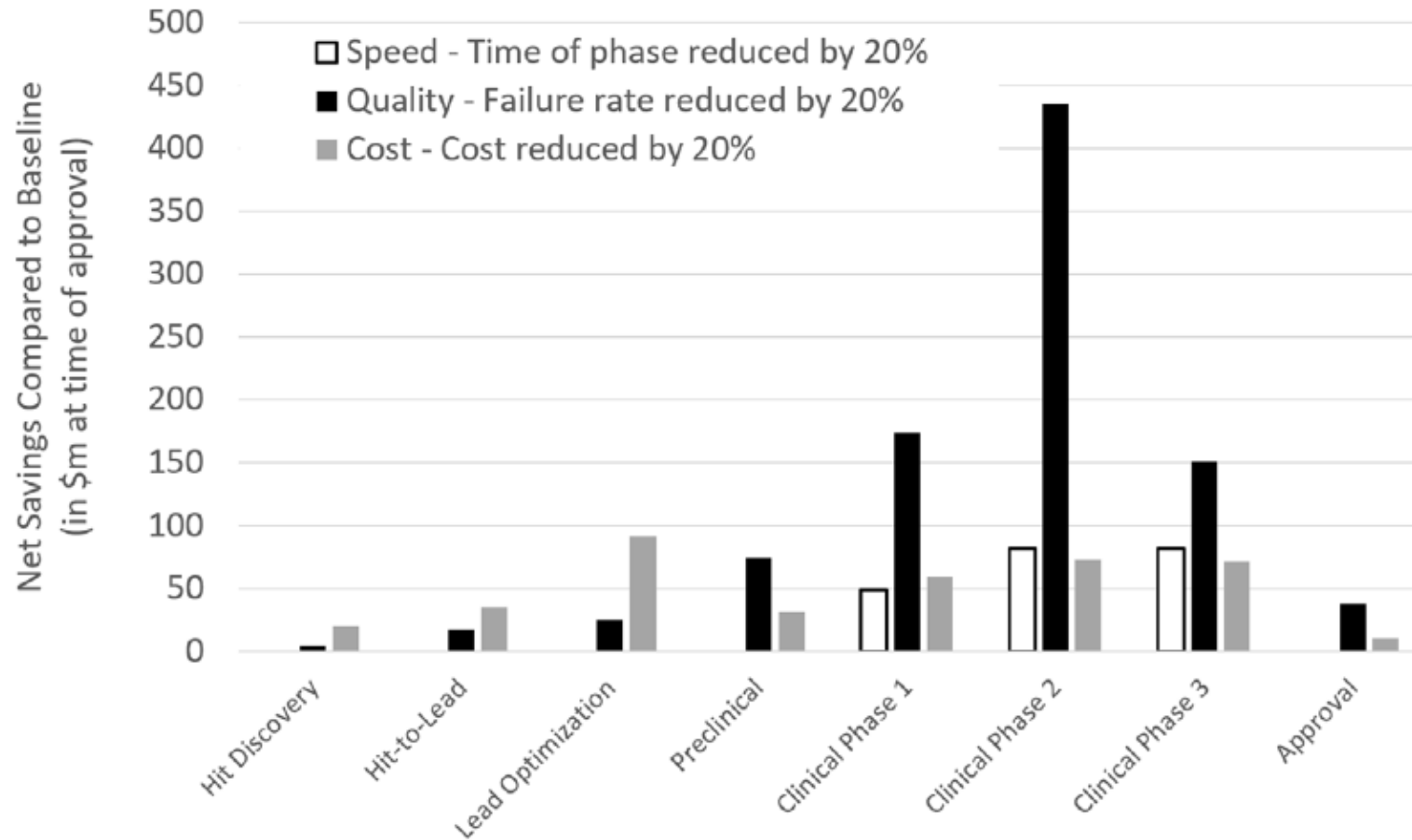
HUN-REN TTK

# A gyógyszerkutatás és fejlesztés folyamata



- Időben hosszú (13.5 év)
- Költséges (873 – 1778 M USD)
- Kockázatos (4% sikerráta)

# Hogyan javíthatunk a folyamaton?



- Költségek csökkentése ?
- Gyorsaság növelése ?
- A gyógyszerjelöltek minőségének javítása, azaz a kockázatok csökkentése !

# A kockázatcsökkentés lehetőségei

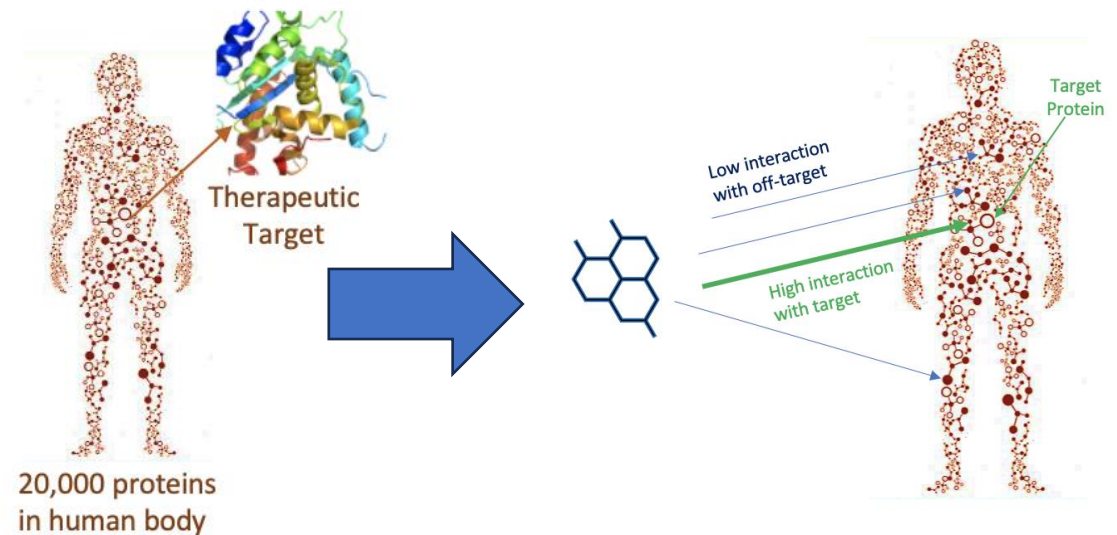
- Jobb minőségű gyógyszerjelöltek (jelenleg 4% sikerráta)
- Jobban validált gyógyszer-célpontok (jelenleg 5-10% validált)
- Jobb betegbeválasztási kritériumok (jelenleg 20-50% beválasztási veszteség)
- Jobb minőségű klinikai vizsgálatok (jelenleg 30% kiesés, 40% megfelelés)



Jobb döntések születnek



Ehhez jobb minőségű adatok kellene



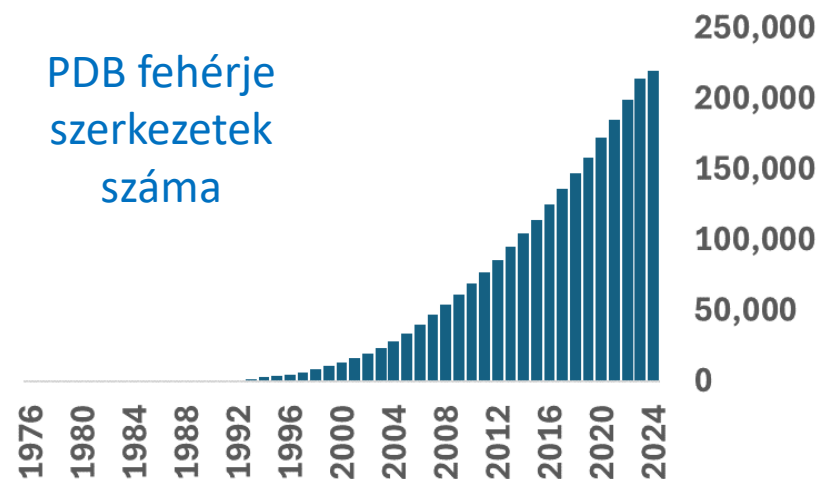
# Az adatok hozzáférhetősége

Adattípus	Adatpontok száma	Adatkészlet mérete (B)
Tesla szenzoradatok	$2 \times 10^7$	$10^{21}$
Gyógyszer szerű molekulák	$10^{63}$	
REAL gyógyszer szerű molekulák	$1.2 \times 10^9$	$10^{15}$
CAS fizikailag előállított molekulák	$2.8 \times 10^8$	
ZINC megvásárolható molekulák	$7.5 \times 10^6$	
ChEMBL biológiai adatok	$2 \times 10^7$	$10^9$
DrugBank forgalmazott gyógyszerek	13,548	
DrugMatrix genetikai adatkészlet	627	
SIDER mellékhatások	1430	
DILI gyógyszer-máj kölcsönhatások	1036	

# Adatbázisok fejlesztése és alkalmazása

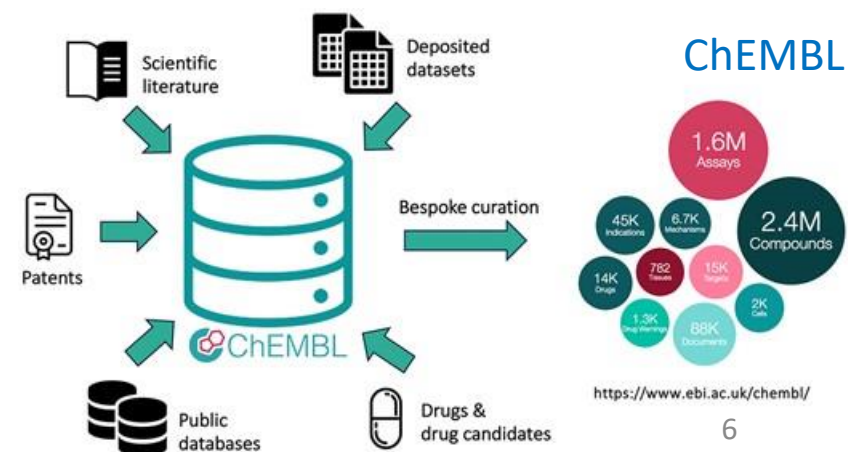
## • Adatbázisok

- PDB 1971: fehérje szerkezetek
  - >200K szerkezet 2024-ben
- CSD 1965: kis molekulák szerkezete
  - ~1300K szerkezet 2024-ben
- PubChem 2004: kémiai és biológiai adatok
  - 118,025K vegyület, 295,078K teszt 2024-ben
- ChEMBL 2008 (korábban StARlite): kémiai és biológiai adatok
  - 2400K vegyület, 1600K teszt 2024-ben



## • Adatbányászat

- Hasonló molekulák keresése
- Tulajdonságterek feltérképezése



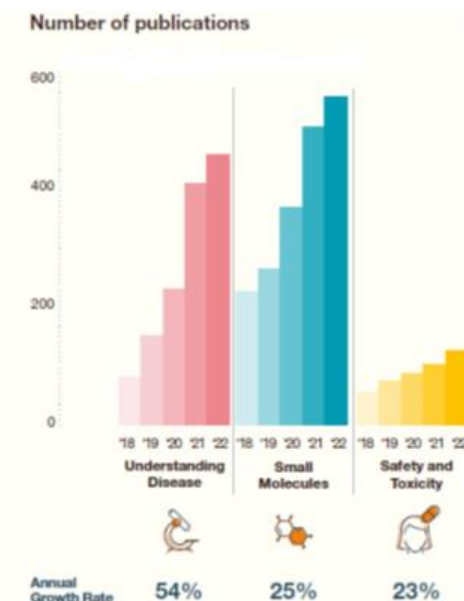
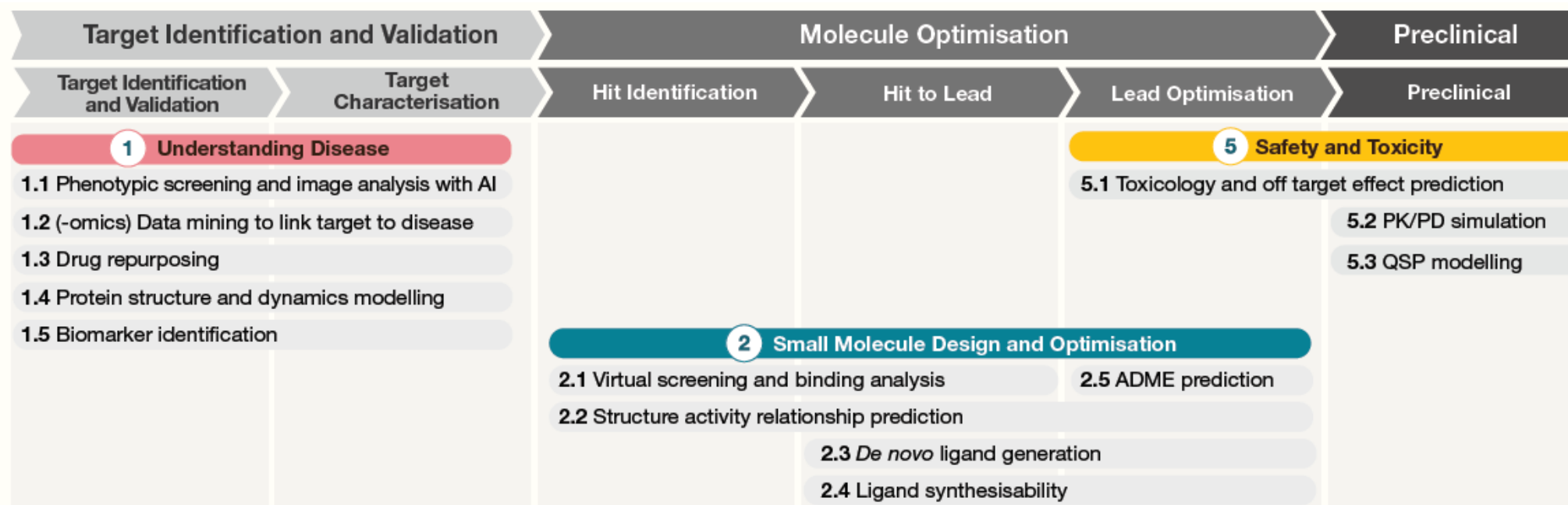
# Preklinikai MI alkalmazások

## Jobb gyógyszer-célpontok azonosítása

- Lehetséges gyógyszer-célpontok előrejelzése
- Potenciális gyógyszer-célpontok validálása

## Jobb gyógyszerjelöltek azonosítása

- Lehetséges ligandumok előrejelzése
- Ligandumok szintézisének javítása
- Ligandumok optimalálása



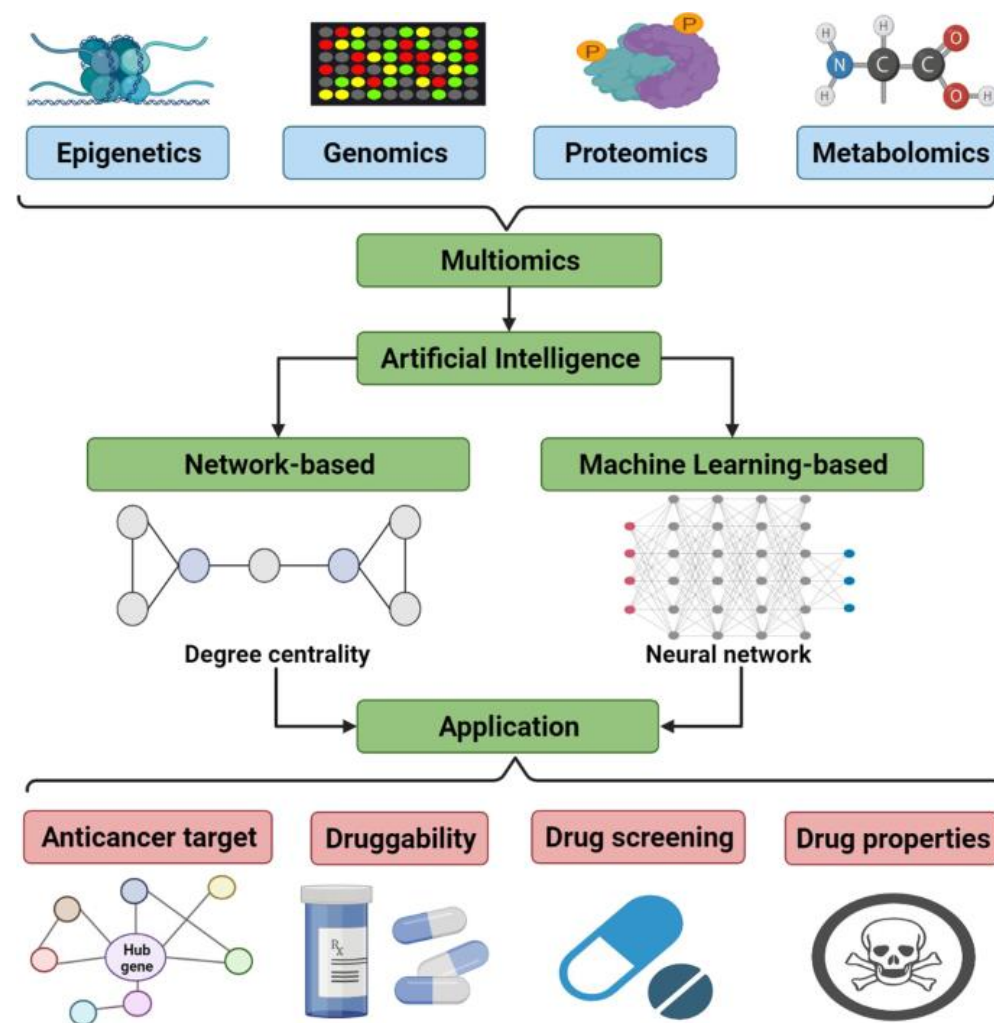
# Gyógyszercélpontok azonosítása és validálása 1.

- Célkitűzések

- Hatásmechanizmus előrejelzése
- Fenotípusos hatás előrejelzése
- In vivo PoC hatás előrejelzése

- Adatforrások

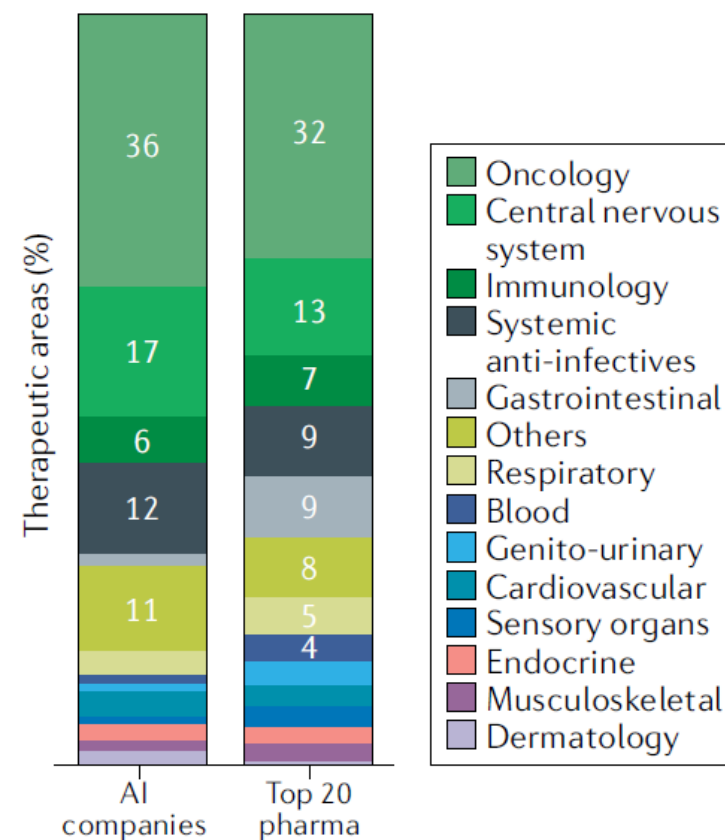
- Genome wide association studies
- Funkcionális genomikai DB
- CRISPR, siRNA, KO adatkészletek
- 3D modellekből származó adatok
- In vivo adatok





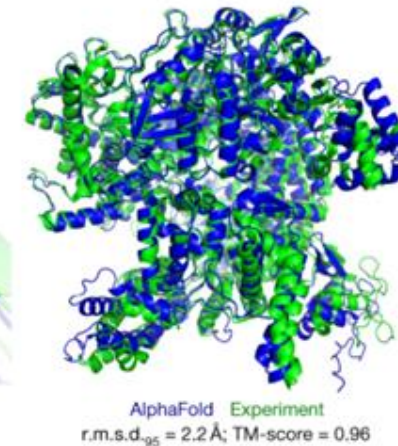
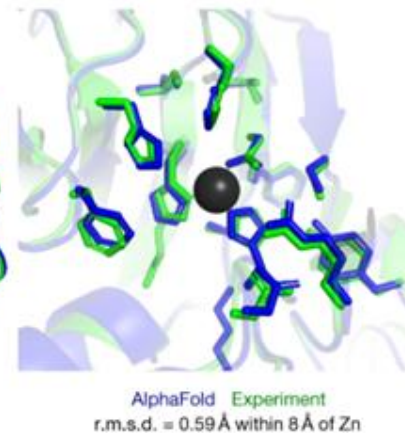
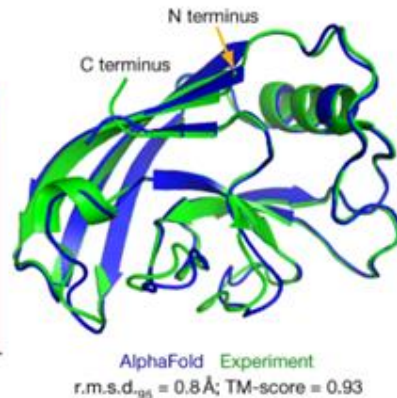
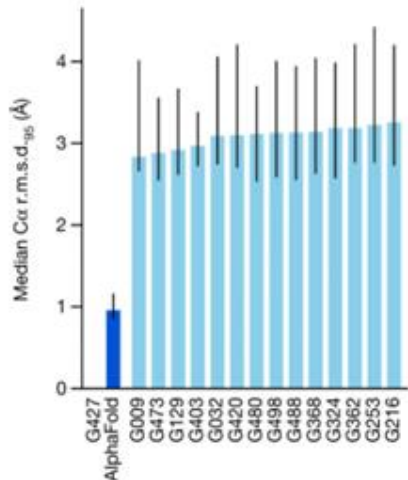
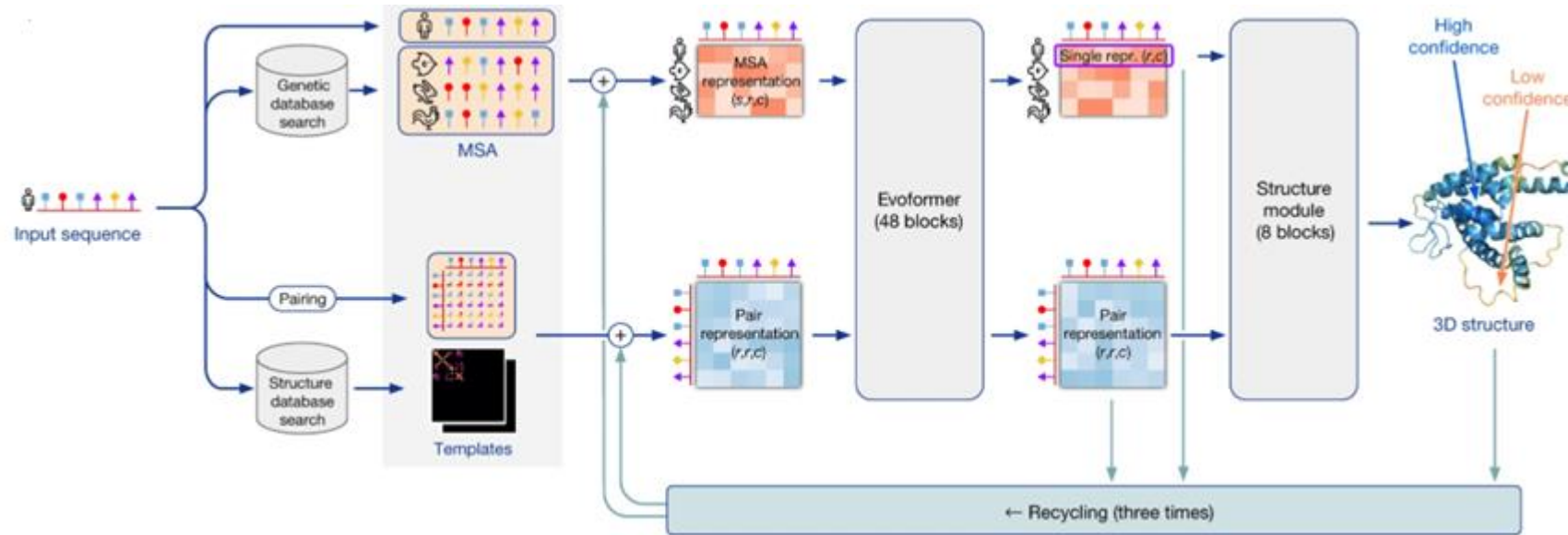
# Gyógyszercélpontok azonosítása és validálása 2.

- Előny
  - Nagyszámú kapcsolt biológiai adat
- Korlátok
  - Splice változatok, aktivációs állapot, sejt típus,...
  - Betegség – fehérje asszociáció érvényessége
- MI hatása a célpont választásra
  - Nincsen változás a domináns terápiás területekben



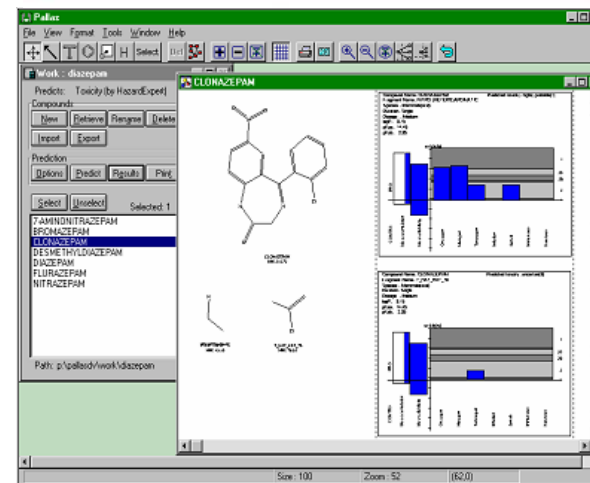
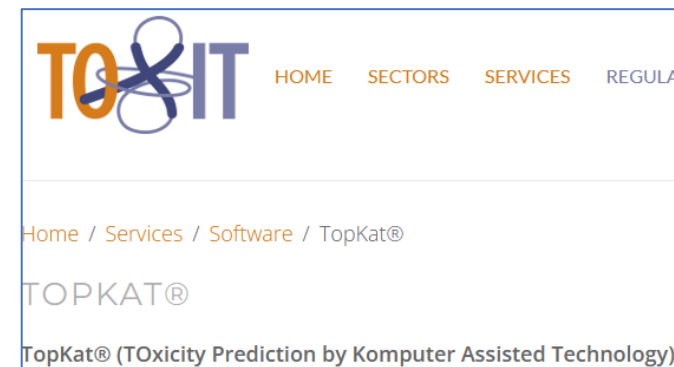
# Gyógyszercélpontok szerkezetének előrejelzése

- AlphaFold – Google DeepMind majd a funkció predikciója



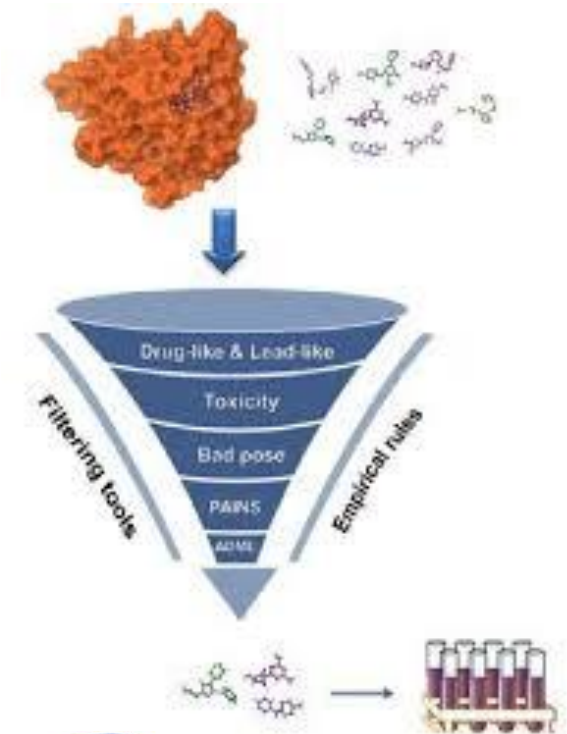
# Gyógyszerjelöltek azonosítása – a kezdetek

- QSAR
  - Biológiai hatás előrejelzés
  - Hansch 1961
    - Biológiai hatás - fizikai-kémiai tulajdonságok
  - Free Wilson 1964
    - Biológiai hatás – szerkezeti paraméterek
- „Szakértői rendszerek”
  - Metabolizmus előrejelzés
    - METEOR
  - Toxicitás előrejelzése
    - DEREK, StAR, TOPKAT
  - CompuDrug '80-as évektől
    - MetabolExpert
    - HazardExpert

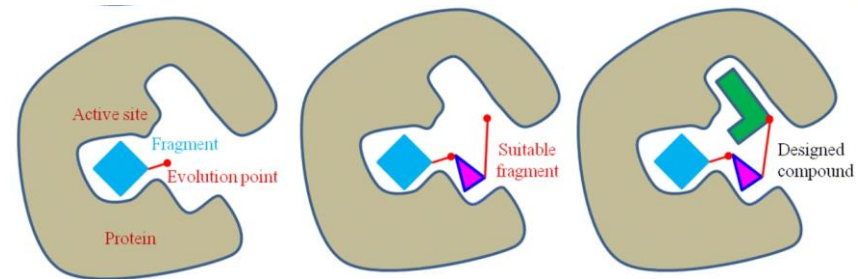


# Gyógyszerjelöltek azonosítása napjainkban

- Virtuális szűrés fehérje szerkezettel: kísérleti vizsgálatra válogatás
  - “hagyományosan:” néhány millió vegyületből
  - MI-vel támogatva: Néhány milliárd vegyületből – kémia tér ( $\sim 10^{63}$ ) lefedettsége nő

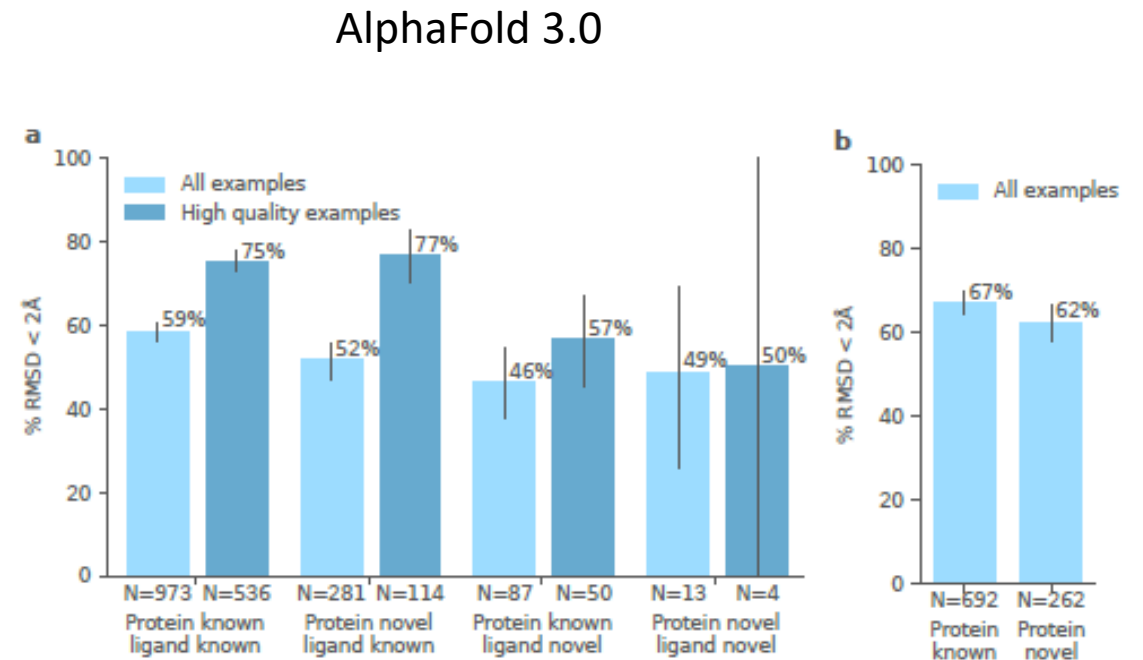
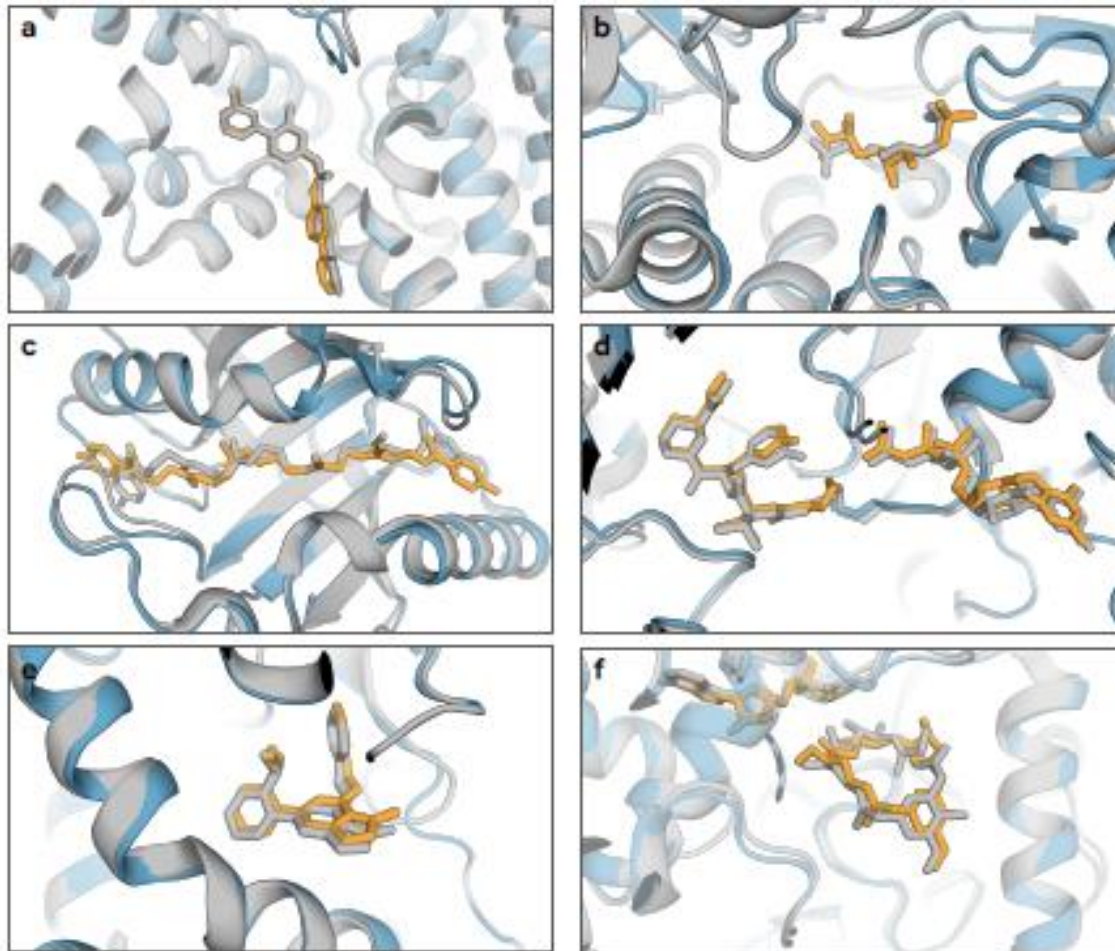


- De novo tervezés
  - Kívánt tulajdonságú vegyület keresése a kémiai térben
    - Kötődés, gyógyszerszerűség, szintetizálhatóság



- „Repurposing”
  - Létező gyógyszer más terápiás célra

# Gyógyszerjelöltek azonosítása – a jövő



# Gyógyszerjelöltek optimalizálása

In vitro modell	In vivo
Aktivitás a célfehérjén	Hatékonyságot farmakokinetika modulálja
Fizikai-kémiai tulajdonságok	Befolyásolja az abszorpciót, metabolizmust, ...
Farmakokinetikai modellek: mikroszomális stabilitás, CYP gátlás,..	Részben tükrözik szövetek és szervek komplex válaszát
Toxicitás: hERG, DILI, Ames,...	Terápiás index

- Az in vitro modellek végpontjainak előrejelzése MI-vel javítható
- Az előrejelzés hasznossága a modellek relevanciájától függ

# Az optimalálás gyakorlata

- Többdimenziós optimalási feladat
- Az ADMET optimalálás összetett folyamat
  - Enzimek, transzporterek, utak, szervek
  - Kor, nem, genetika, egészségi állapot
- Adatok:
  - in vivo - kevés
  - In vitro – releváns modell?
  - Feltételes (dózis függő, ...)
  - Nem homogén
  - Részben publikus
  - Lokális modellek állíthatók fel
  - Eloszlás ismeretlen, extrapolálás bizonytalan
- Validálás, előrejelzés kérdéses

## Optimálendő tulajdonságok

Hatás

Mellékhatás

Felszívódás (A)

Eloszlás (D)

Metabolizmus (M)

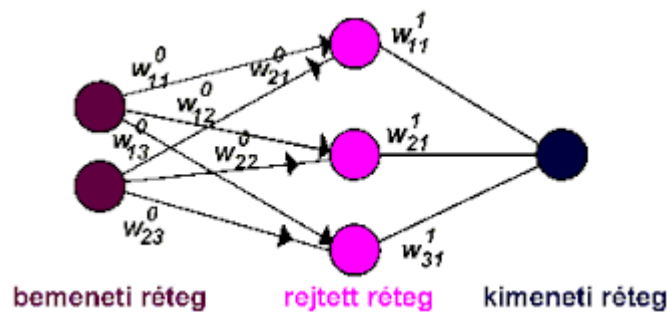
Kiürülés (E)

Toxicitás (T)

...

# Példa: Központi idegrendszeri aktivitás előrejelzés

- Központi idegrendszeri (CNS) aktivitás – vér- agy gát áthaladás
- Előrecsatolt neurális hálózat (992x31x1) a CNS aktív vegyületek azonosítására
- Molekula reprezentáció: ujjlenyomat
- Molekula hasonlóság alapú előrejelzés nem hatékony
- Tanító készlet: 3500 molekula
- Teszt készlet: 3500 molekula
- Tanítás
  - hiba-visszaterjesztés
  - 2000 ciklus
  - Tanítási ráta: 0,2

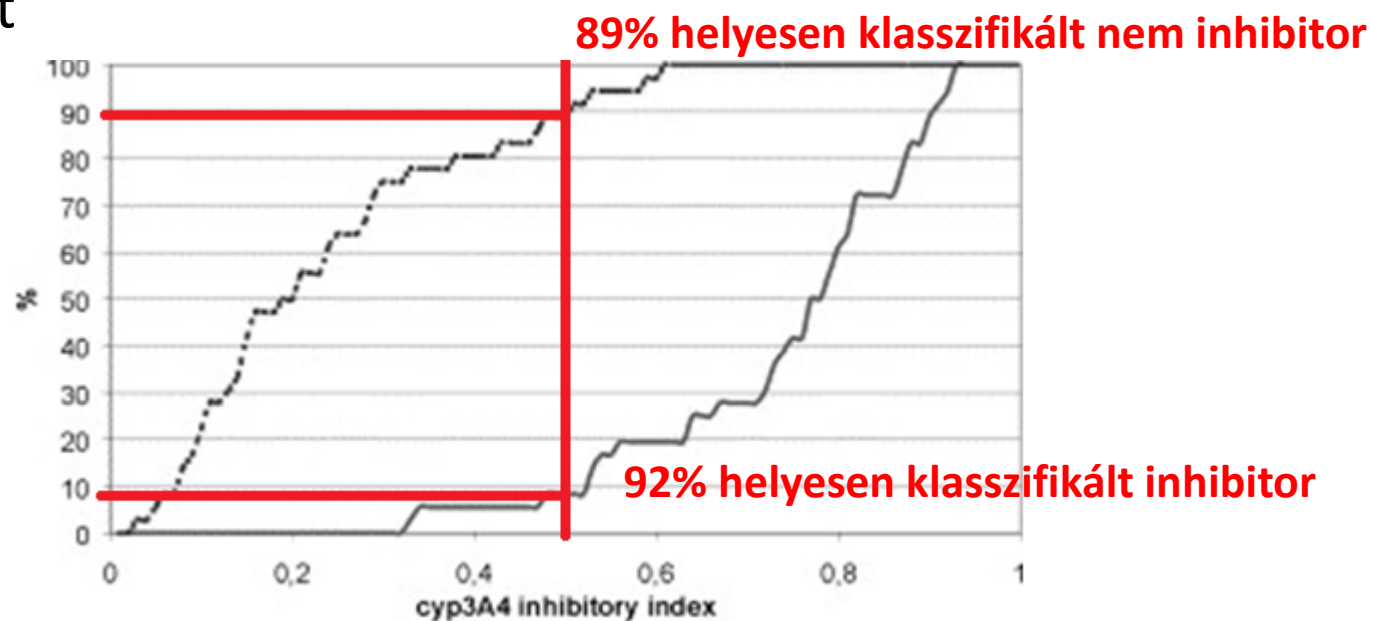


CNS aktivitás	Előrejelző képesség (%)
Antidepresszáns	87,5
Hányáscsillapító	81,3
Antipszichotikum	90,4
Anxyolitikum	81,6
Dopaminerg	98,2
Muskarinerg	96,2
Szerotoninerg	93,0
<b>Összesen</b>	<b>89,9</b>



# Példa: Citokróm P450 inhibitor azonosítás

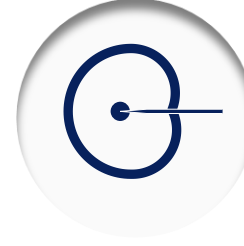
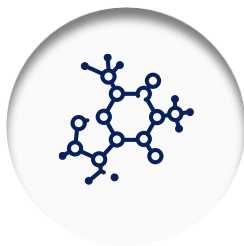
- Citokróm P450 3A4 inhibitor – gyógyszer-gyógyszer kölcsönhatásért felelős
- Előrecsatolt neurális hálózat (992x31x1) az inhibitorok azonosítására
- Molekula reprezentáció: ujjlenyomat
- Tanító készlet: 109 molekula
- Teszt készlet: 36 molekula
- Tanítás
  - hiba-visszaterjesztés
  - 2000 ciklus
  - Tanítási ráta: 0,2



- 92% helyesen klasszifikált inhibitor
- 89% helyesen klasszifikált nem inhibitor

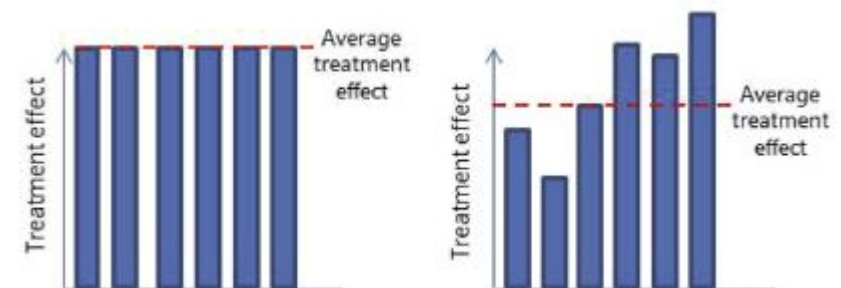
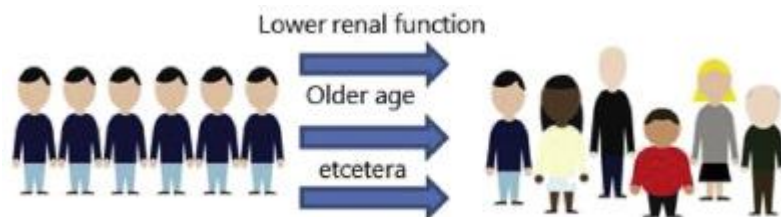
# A preklinikai MI módszerek korlátai

- A ligandumok lehetséges kölcsönhatását nem csak a gyógyszercélponttal, hanem minden lehetséges kölcsönható partnerrel kellene előre jelezni (fehérje, RNS, DNS stb.)
- A gyógyszercélpont expressziójának lokalizációját és esetleges változását is figyelembe kellene venni
- A ligandumoknak el kell jutniuk a hatás helyére, ezért in vivo PK előrejelzés is szükséges lenne

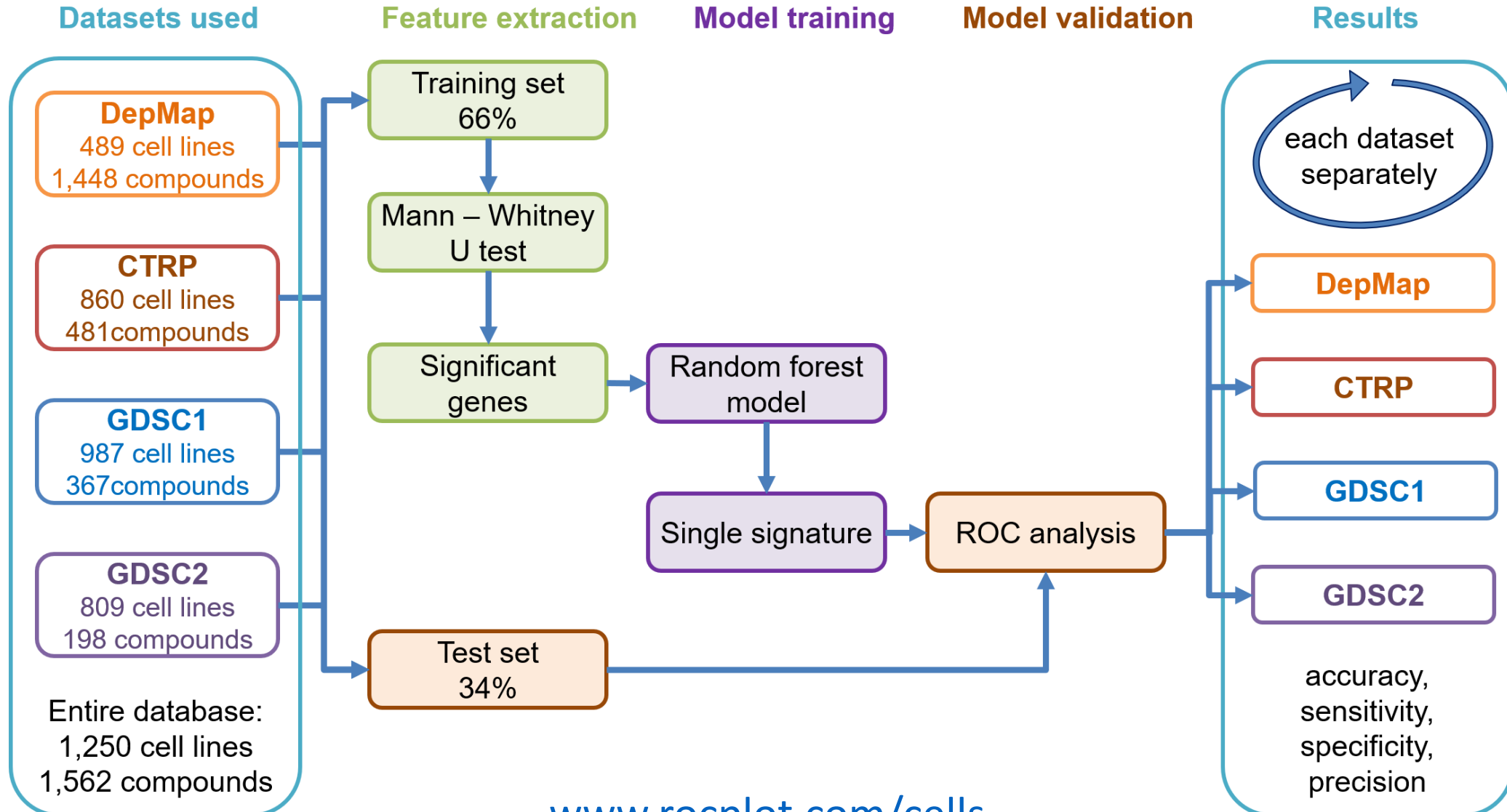


# Klinikai AI alkalmazások

- Célkitűzések
  - Jobb betegbeválasztási kritériumok
  - Jobb minőségű klinikai vizsgálatok
- A rendelkezésre álló klinikai adatok mennyisége nem elégséges
- A klinikai adatok hozzáférhetősége sokszor nem biztosított
- A klinikai adatok megbízhatósága és reprodukálhatósága sokszor kérdéses
- A valós kezelési körülmények között keletkező adatok gyűjtése nem megoldott, minősége, megbízhatósága, standardizáltsága és struktúráltsága kérdéses



# Gépi tanulás génlistákkal: melyik tumor lesz érzékeny a kezelésre

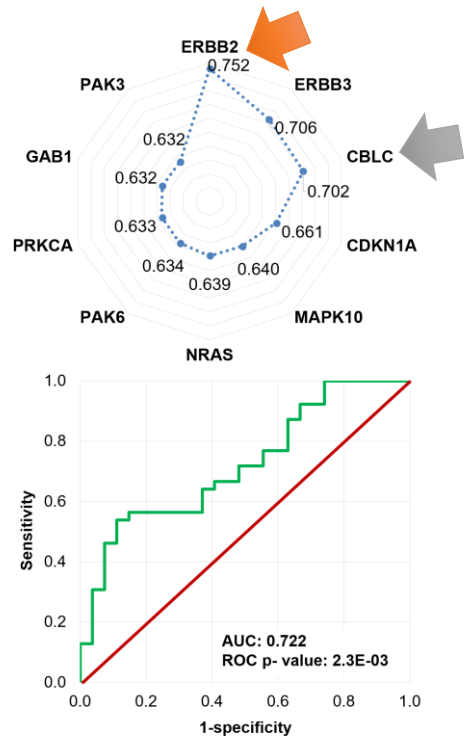


[www.rocplot.com/cells](http://www.rocplot.com/cells)

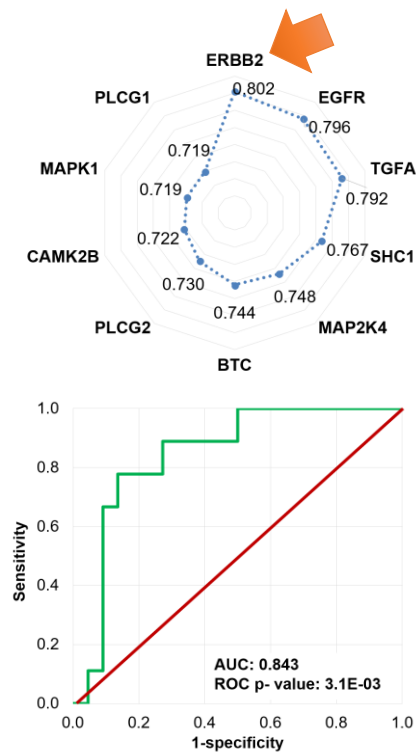
# Érzékenység Trastuzumab kezelésre – ERBB2



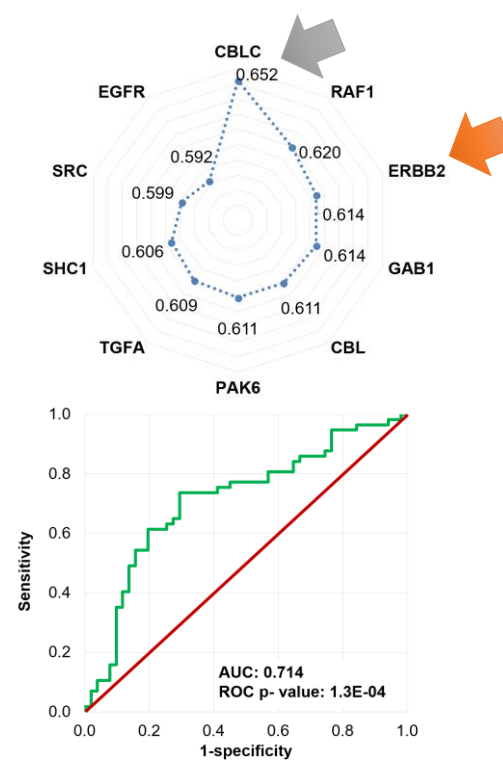
## DepMap



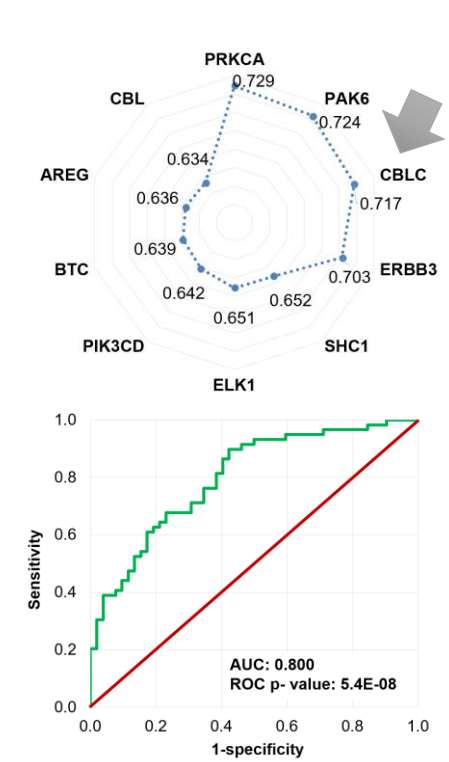
## GDSC1



## GDSC2



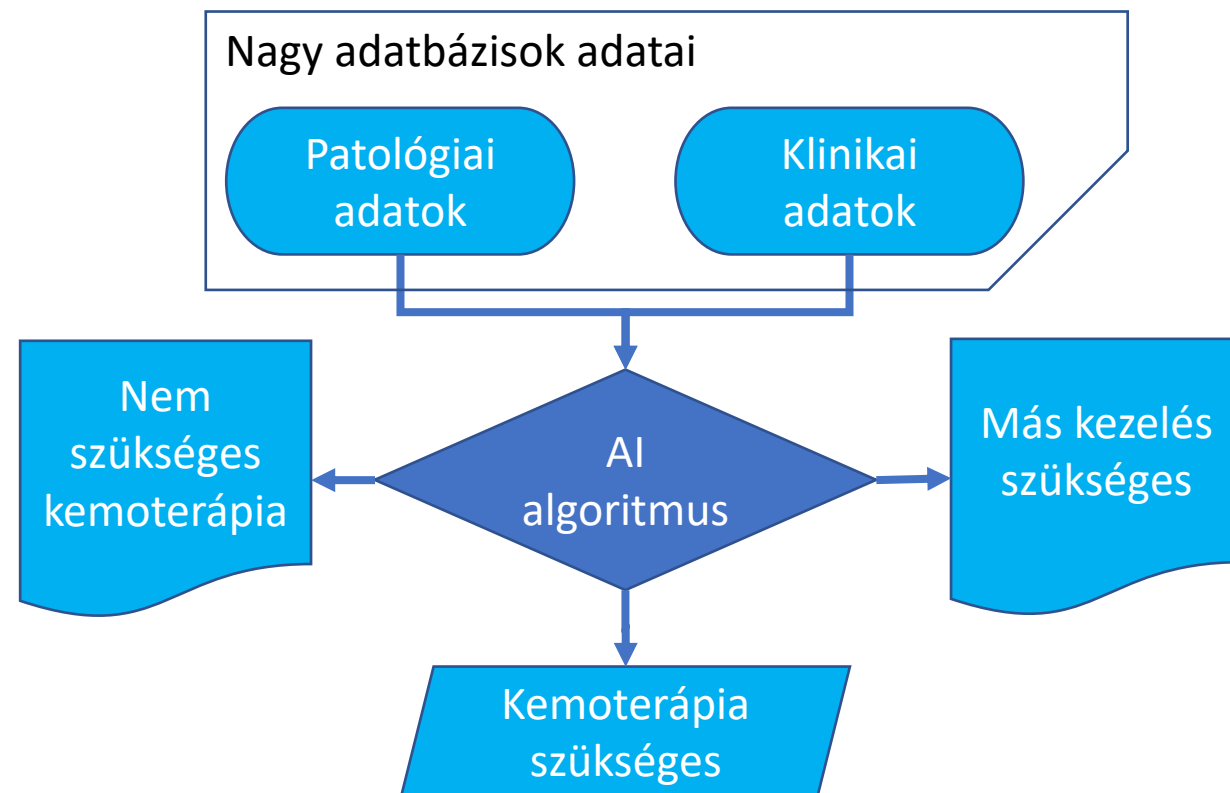
## CTRP



# Leghatékonyabb kezelés kiválasztása



**Cél:** Egy olyan MI modell felállítása, amely emlőtumor esetén személyre szabott túlélési és kezelési becslést tud adni ismert **klinikai és patológiai adatok** alapján.



# ADATBÁZIS



SEER adatbázis  
(2000 utáni betegek)



Nyers adat  
Változók: 39

Sor: **1,154,395**



Feldolgozott adatok

Változók :21

Beteg: **1,026,409**

Kor < 18  
N= 53

Túlélés  
hossz= 0  
N=14,948

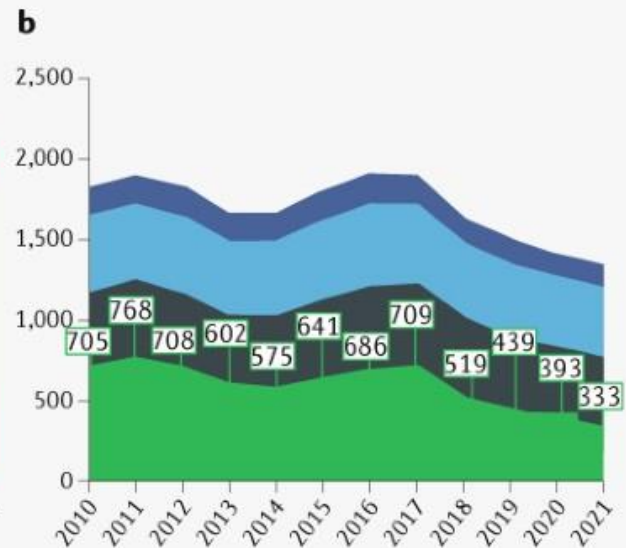
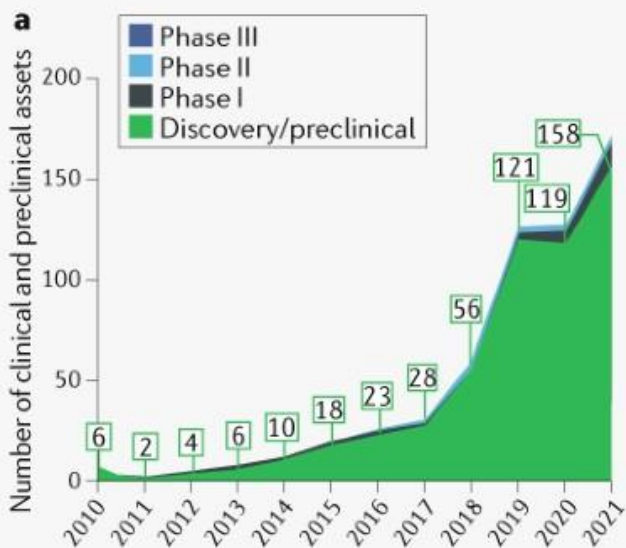
Túlélés flag,  
hiányos dátumok  
N= 60,988

Duplikátumok  
N= 59,369

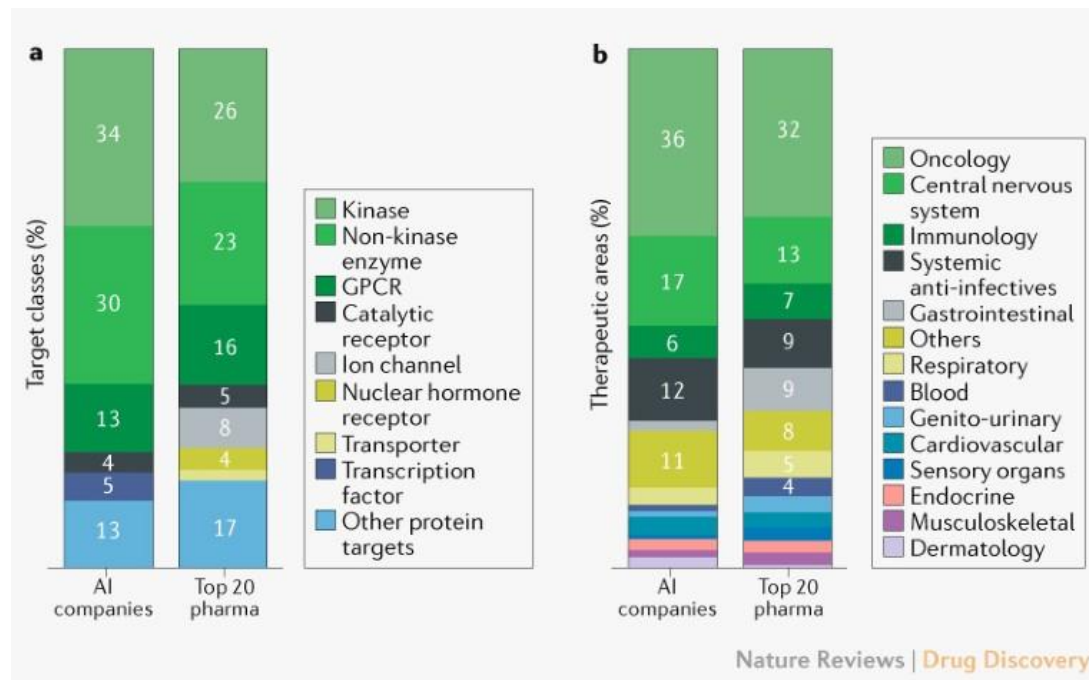
kizárva

Kombinált változók	Eredeti változók
STAGE	4
SIZE	3
GRADE	3
T	4
M	4
5	18

# Mit mutat a valóság?



Nature Reviews | Drug Discovery



Nature Reviews | Drug Discovery

- Az MI módszerek jelenlegi alkalmazása gyógyszerfejlesztésben
  - Elsősorban a preklinikai szakaszra korlátozódik
  - A legtöbb esetben onkológiai és immunológiai indikációkban ismertebb célpontokon
- MI alapon fejlesztett gyógyszerek a klinikán
- Mi alapon fejlesztett gyógyszer nincsen a piacon



# Összefoglalás

- A ma AI-nak nevezett eszközök (gépi tanulás, mélytanulás, mesterséges ideghálózatok, stb.) már évtizedek óta rendelkezésre állnak, ezeket jelenleg is óriáscégek fejlesztik
- Az érdemi használathoz nagy mennyiségű és jó minőségű adatra van szükség
- Ez jelenleg csak a gyógyszerfejlesztés korai stádiumában áll rendelkezésre
  - Nagyszámú adat in vitro modellekre
  - MI eszközökkel az in vitro modellek végpontjainak előrejelzése javítható
- Ezért helyesebb ligandum fejlesztésről mintsem gyógyszerfejlesztésről beszélni



# Összefoglalás

- Ugyanakkor viszont a fejlődés iránya egyértelmű: az AI alkalmazások előtt nyitva az út, amint megfelelő mennyiségű és minőségű hozzáférhető adat lesz
  - Preklinika: Releváns végpontok és kísérleti modellek alkalmazása
  - Klinika: EESZT, Nemzeti Rákregiszter, ELIXIR nemzetközi adatbázisfejlesztések
- Ezért ma a legfontosabb feladat az adatforrások konszolidációja

