

Antitest repertoár jellemzése: egy példán keresztül

p300 és CBP

Antitestek..., de miért is?

Két érdekes fehérje szerkezeti és funkcionális tanulmányozására...

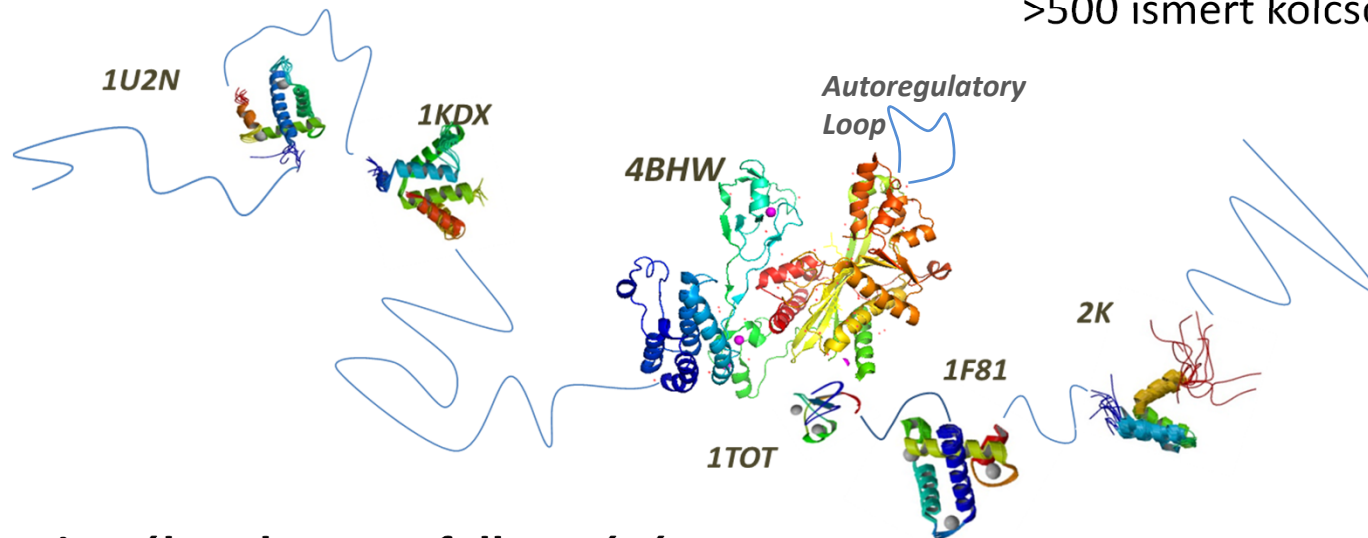
transzkripciós ko-aktivátorok

hiszton acetyltransferáz aktivitás

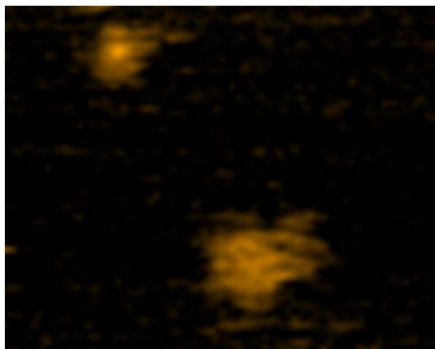
9 rendezett domén és >50% rendezetlen régiók

>150 poszt-transzlációs módosítás,

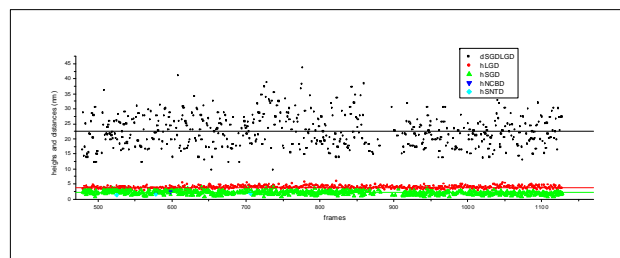
>500 ismert kölcsönható partner



Szerkezetvizsgálat alacsony felbontású egymolekulás módszerekkel...



HS-AFM

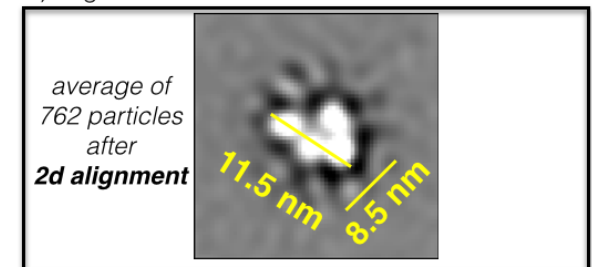


Nagy globuláris domén (LGD) körül mozog egy kicsi (SGD). A kettő közti távolság frame-enként mérhető ~22 nm

A dinamikus SGD-t az átlagolás során elveszítjük. Specifikus jelölés kéne!

EM

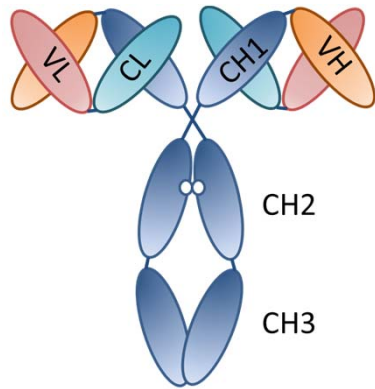
3) Alignment 2D:



Specifikus jelölés specifikus antitestekkel

Állat immunizálása az antigénnel

→ specifikus antitesteket kódoló immunsejtek szelektálódnak a természetes immunválasz során (szomatikus hipermutáció)



Hagyományos antitest ~150 kDa

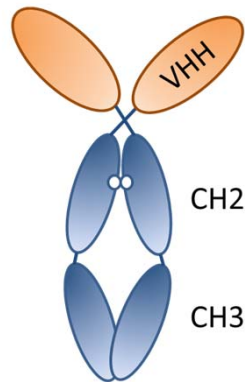
Antiszérum

Tisztított poliklonális antitest

} Főleg detektálásra, izolálásra alkalmas

Monoklonális antitest → molekuláris jelölésre is alkalmas

Probléma: nagy és 4 alegység komplexe



Nehézlánc antitest ~100 kDa

Izolált limfocitákból RNS-t tisztítanak
Reverz transzkripcióval cDNS-sé írják a VHH régiókat
PCR, klónozás fág vektorba

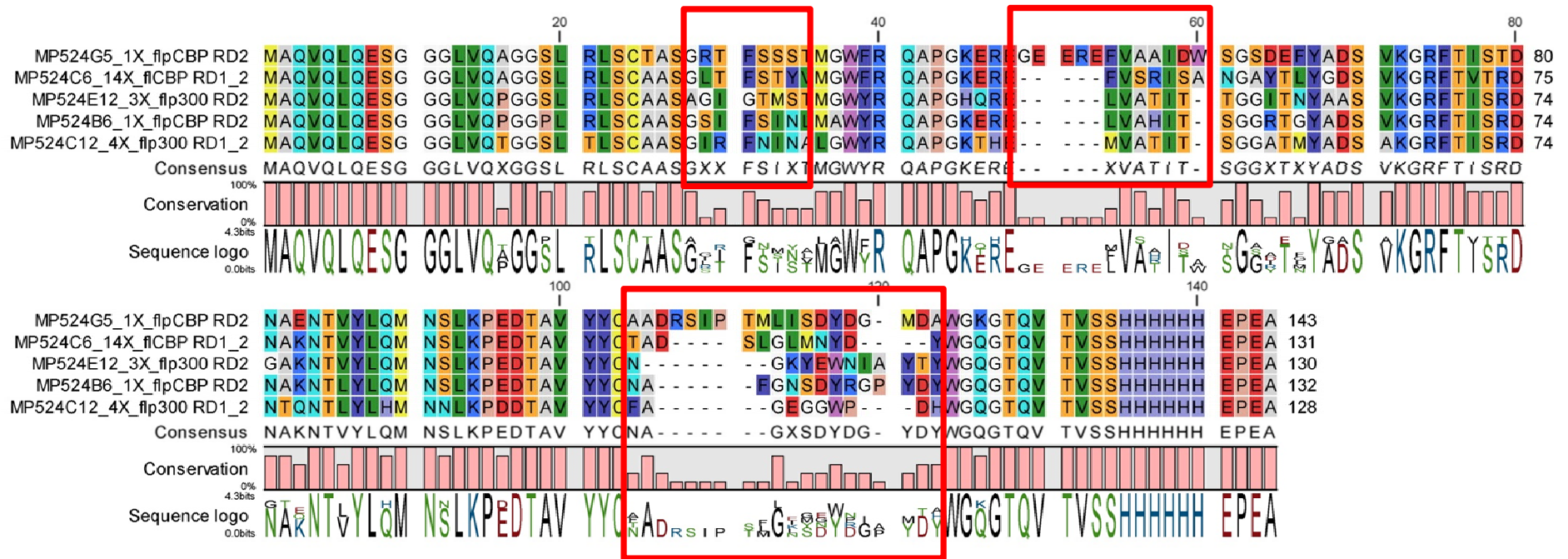
→ Fág bemutatás (phage display) → **szelektálás**



Nanobody ~15 kDa

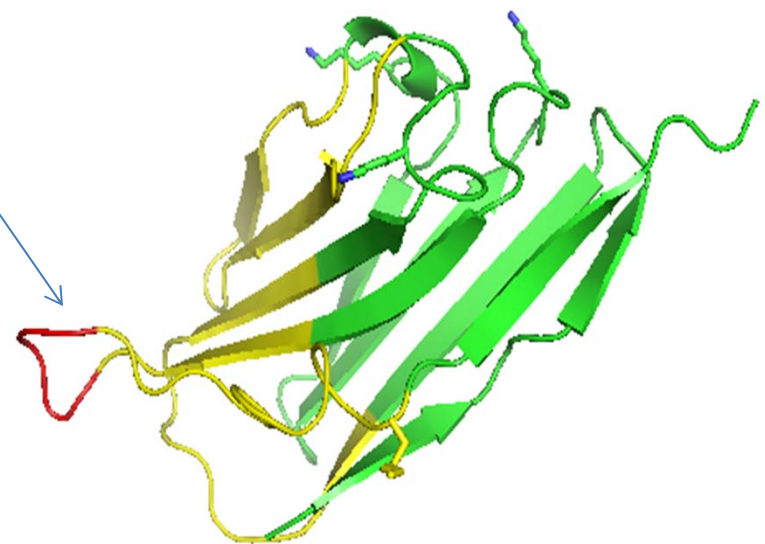
Fagemiddel transzformált baktériumok periplazmájában expresszálható → **clonal screening**

Nanobody scaffold, CDRs, structure

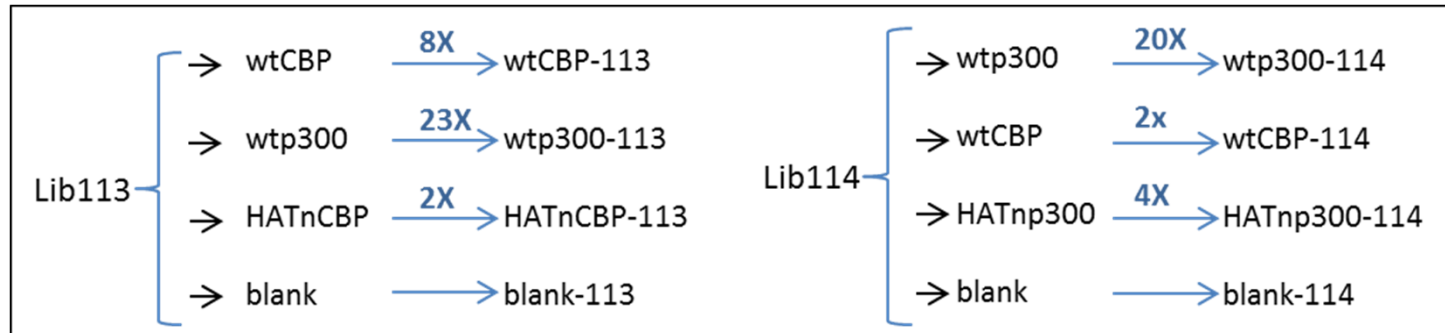


CDR3

Modell szerkezet az egyik izolált és részletesen jellemzett, CBP KIX doménra specifikus nanobody-ra.
 A CDR3 régióban azonosítható a jellegzetes KIX kötő motívum is ($\Phi xx\Phi\Phi$, pirossal jelölve a szerkezetben).



Szisztematikus szelektálás → az alkönyvtárak jellemzése NGS-sel



8 alkönyvtár szekvenálása:

8 féle szekvenciakód ligálása a 8 mintához (index), adapter ligálás

Szekvenálási könyvtár készítése PCR-rel (1 szekvenálásban a 8 minta)

Szekvenálás: 2X300 bp read-ek, Illumina MiSeq platform

Cél:

- **teljes lefedettség** a nanobody-t kódoló régiókra
- Minél **mélyebb szekvenálás** → minél több klón elolvasása
- Az egyes klónok **előfordulási gyakoriságának** mérése a különböző mintákban

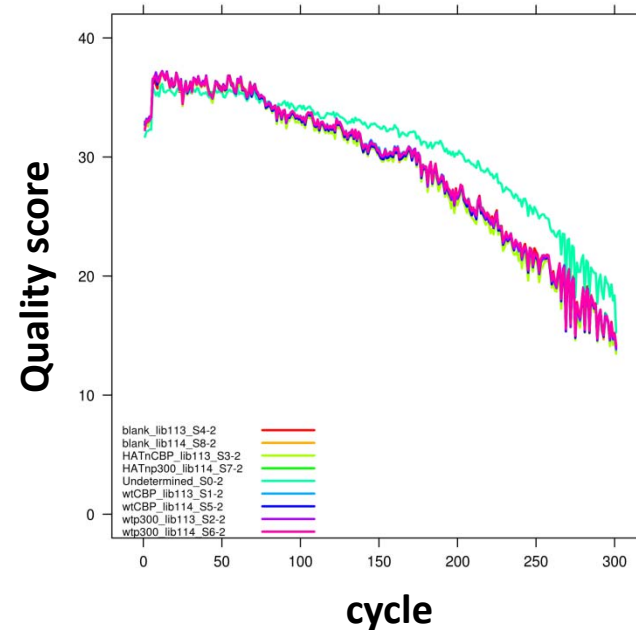
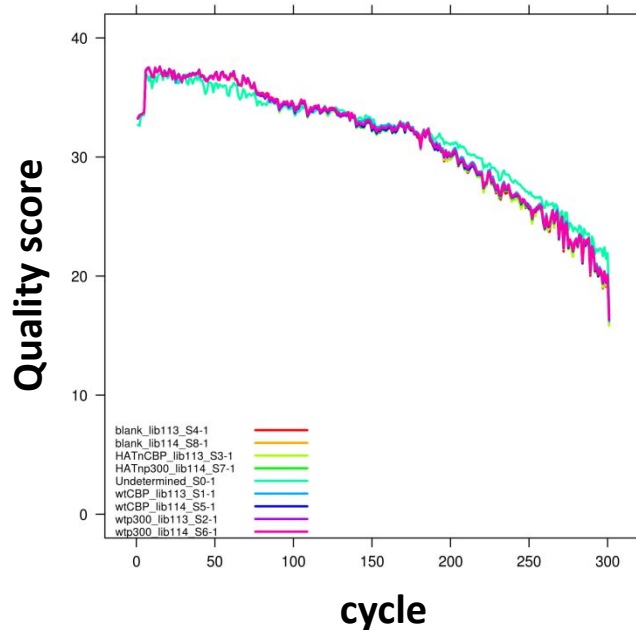
Problémák:

- A nagyon hasonló szekvenciák miatt okozott PCR egyensúlytalanság kiküszöbölésére spike-in DNS-t is keverték a mintákhoz
- Szekvenálási/PCR hibák elkülöníthetlensége a szomatikus hipermutáció okozta változatoktól
- A rokon változatok klaszterezése
- A frekvencia adatok összehasonlítása és érthető ábrázolása

Eredmények összesen 11,3 Gb, 18,85 millió raw read a 8 könyvtárra és a spikein-ra együtt
→ quality filter: a readok ~80%-a Q20-as vagy jobb

$$P(\text{error}) = 10^{-(Q_{\text{score}}/10)}$$

99% biztonsággal hívott bázisok

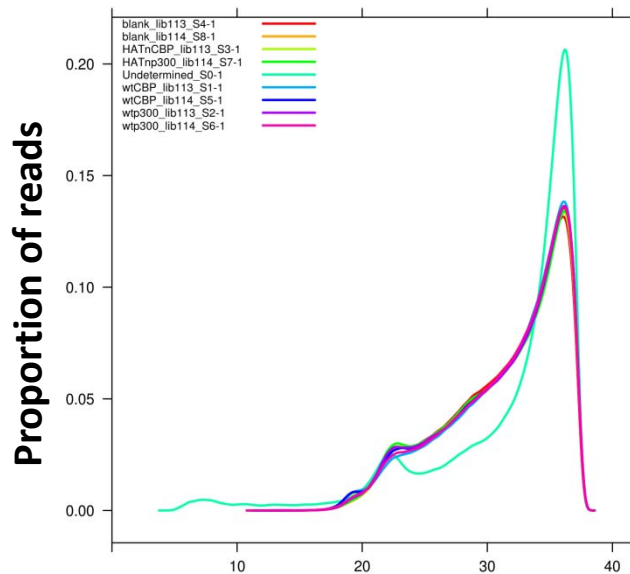


- trimming: az alacsony Qscore-ú régiók és az esetleges adapter szekvenciák levágása, 35 b-nél rövidebb read-ek eliminálása (~99,5% maradt), további minőségi szűrés: ~90% maradt
- Stitching: a minőségi szűrők miatt előfordul, hogy a read párok egyike esik csak ki...
~10% forward, és ~4% reverz read volt ilyen „broken pair”
- 13,1 millió fragmens (nanobody) DNS szekvencia szintjén a 8 könyvtárra együtt
- 559 988 egyedi aminosav szekvencia...

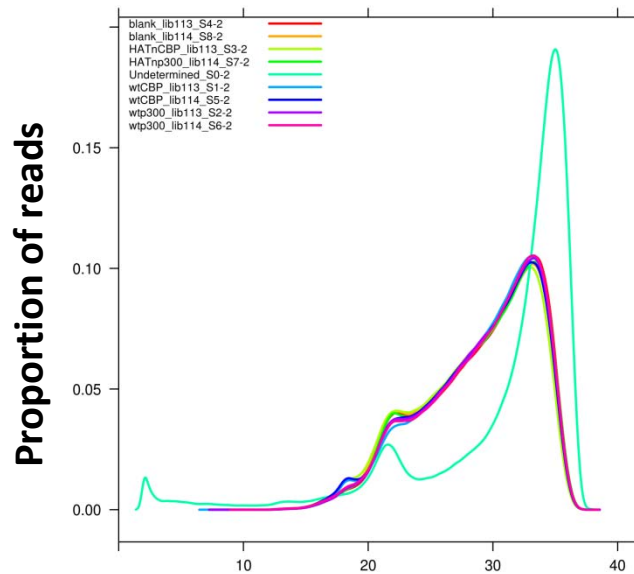
Eredmények összesen 11,3 Gb, 18,85 millió raw read a 8 könyvtárra és a spikein-ra együtt
→ quality filter: a readok ~80%-a Q20-as vagy jobb

$$P(\text{error}) = 10^{-(Q_{\text{score}}/10)}$$

99% biztonsággal hívott bázisok



Average base quality



Average base quality

- trimming: az alacsony Qscore-ú régiók és az esetleges adapter szekvenciák levágása, 35 b-nél rövidebb read-ek eliminálása (~99,5% maradt), további minőségi szűrés: ~90% maradt
- Stitching: a minőségi szűrők miatt előfordul, hogy a read párok egyike esik csak ki...
~10% forward, és ~4% reverz read volt ilyen „broken pair”
- 13,1 millió fragmens (nanobody) DNS szekvencia szintjén a 8 könyvtárra együtt
- 559 988 egyedi aminosav szekvencia...

559 988 egyedi aminosav szekvencia: → túl nagy és a szekvenálási hibák miatt redundáns adatszett...

* 1aa eltérés miatt már külön kezeli a szekvenciákat, így a vonatkozó frekvencia adatok is szétaprózódnak...

* a kalszterezés nem volt kivitelezhető... az összes szekvencia összes másik szekvenciával való összehasonlítása kb. 360 milliárd páronkénti összehasonlítást jelenten

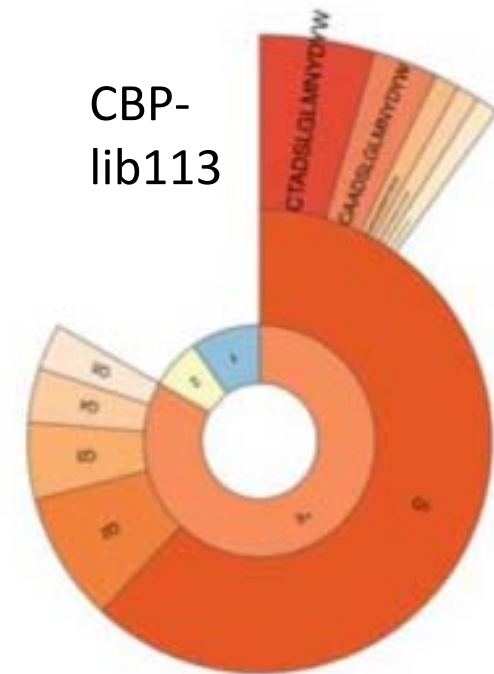
Clonotype profiling

Klonotípusokat definiáltunk a CDR3 régió alapján

→ Összesen 355 023 klonotípus lett

MIXCR v2.1 / VDJtools

- VDJ alignment: a párosított előprocesszált read-ek Referenciához (humán(!) V/D/J/C gének) való illesztése
- Assemble: a CDR3 régiók alapján összeállítja az egyedi klonotípusokat
- ExportClones: out-of-frame és stop kodonos klonotípusok kiszűrése
- JoinSamples: a 8 mintára vonatkozó klonotípusok összevetése, frekvencia adatokkal (a klonotípushoz illesztett readek száma / összes readek száma a mintában) → xls file
- Annotate: a CDR3 szekvencia jellegzetességei alapján prediktál különböző tulajdonságokat...

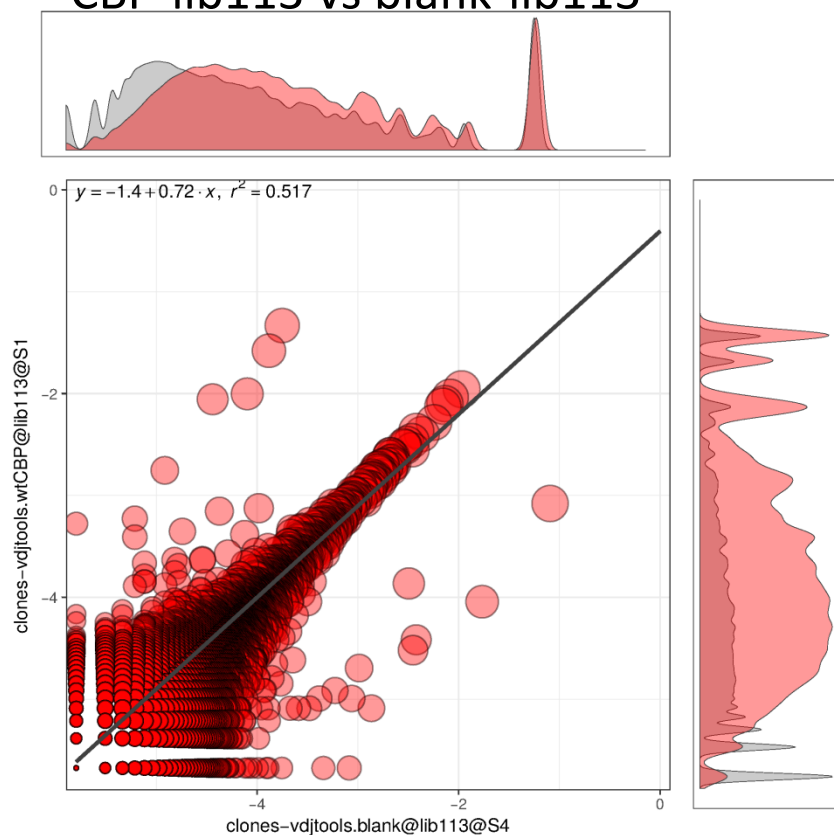


Eredmények ábrázolása frekvencia korrelációs grafikonon

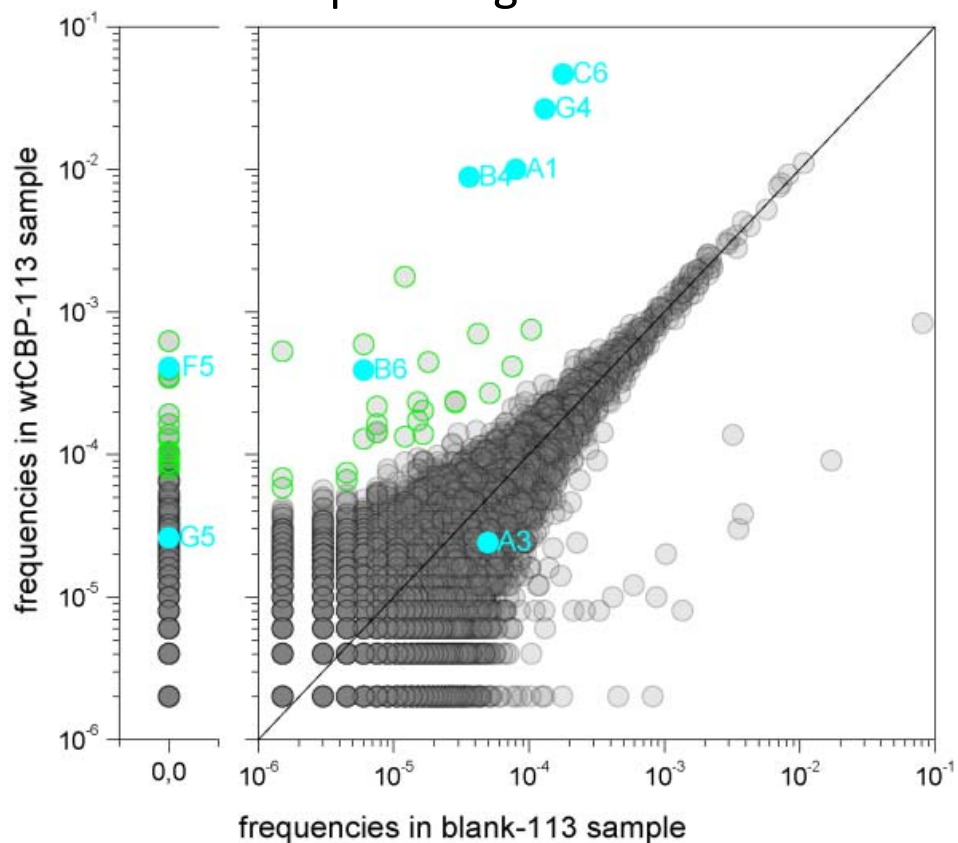
vdjtools OverlapPair

Itt csak a két mintában közös klonotípusok szerepelnek

CBP-lib113 vs blank-lib113



Nekünk az is érdekes, ami a kihúzottban megvan, de a blank-ben nem jött ki...
Módosított plot Origin 8.6-ban



Összesen 162 ígéretes jelölt

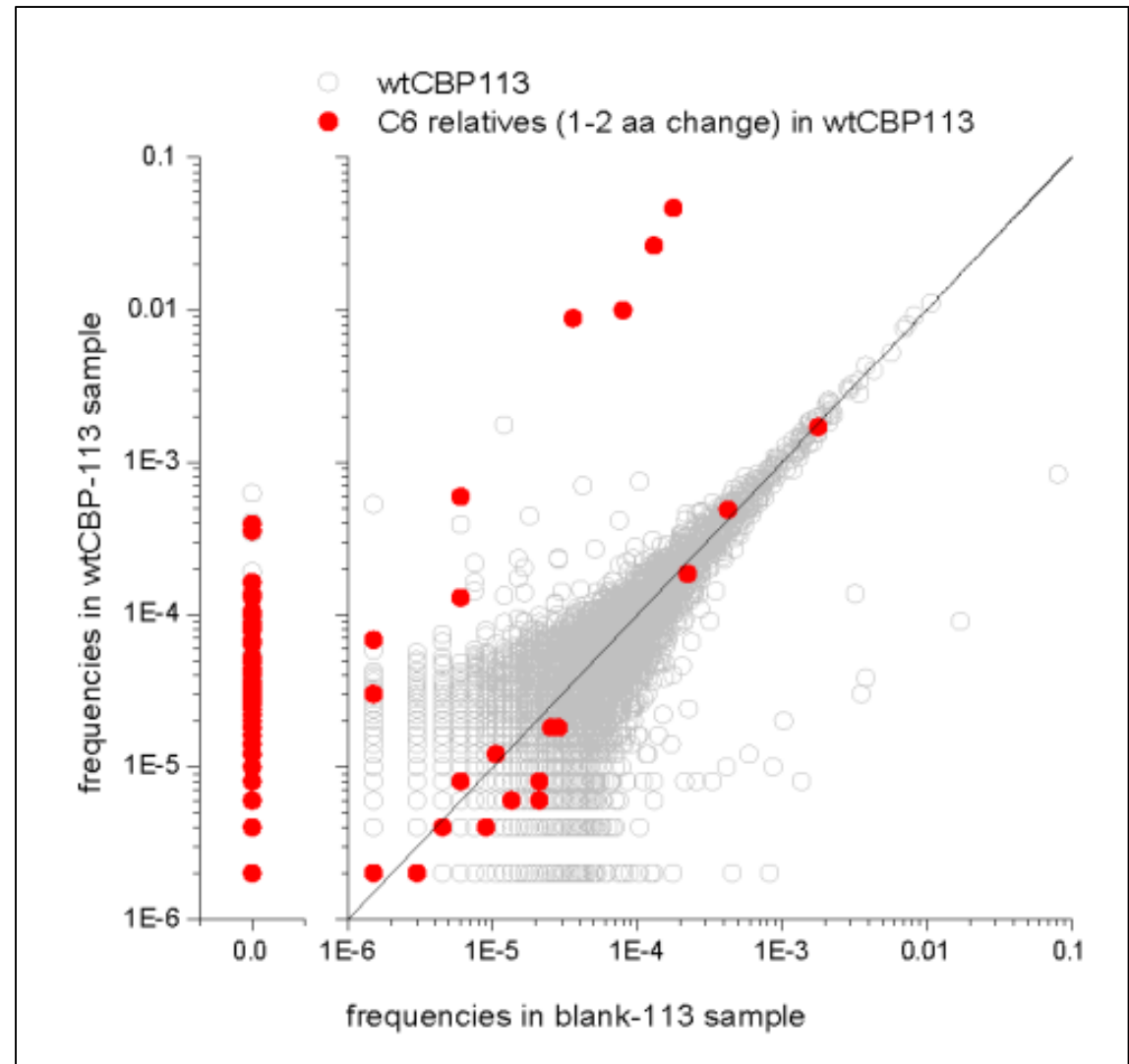
- 91 egyedi klonotípus
- 71 klonotípus összesen 17 családba sorolható

Ezen a grafikonon kb. 80ezer adatpont van
Korábbi kísérletekben azonosítottak...
Ígéretes jelöltek

Néhány karakterizált nanobody-ra megnéztem, van-e rokona.

A CDR3 szekvenciákból egy helyi blast adatbázis készítettem, és blast-tal kerestem a kiszemeltekhez hasonlókat

A C6 egy népes családhoz tartozik, aminek több tagja az erős jelöltek közé tartozik.



A fenti anyag egy esttanulmány, aminek a **specifikus részleteit nem kell megtanulni!**

Ami **kell** belőle:

- 1) Megérteni az antitestek mibenlétét, és keletkezésük módját (szomatikus hipermutáció és mikroevolúció)
- 2) Az antitestek variábilis régióinak, pl. a nanobody-knak az alapvető felépítése (CDRs)
- 3) Az NGS-sel történő antitest repertoár jellemzés főbb kihívásai (teljes lefedettség a variábilis régióra, frekvencia számolás, szekvenálási/PCR hibák elkülönítése a SHM okozta változatoktól... egy ilyen antitest repertoár változatosságának átgondolása)
- 4) Az NGS-es adatok értékelésének lehetőségei, néhány alapfogalom ismerete (quality check, Q20 értelmezése, trimming, stitching, klonotipizálás, overlap pair plot értelmezése, eszközök, mint a VDJ Tools...)