

Neyroinformatika - genlərdən hüceyrələrə və sinir sistemində

ANS Summer School

21.08.2025

Ərtoğrul Alışbəyli

1. Giriş

Gəlin müzakirəmizin əsasını qoymaq adına bir sualla başlayaq: Neyroinformatika dedikdə ağılınıza nələr gəlir?

{İnteraktiv müzakirə}

Neyroinformatika təməldə müasir hesablama imkanlarından istifadə edərək bioloji datanın işlənməsi üçün metodlar istehsal edib bu metodları sinir sistemini anlamaq üçün tətbiq edən sahədir. Bir tərəfdən baxanda neyroinformatikada nəzəri olaraq həqiqətən də yeni adlandırma biləcəyimiz şeylər baş vermir: Sadiqin də izah etdiyi kimi, müasir elmi metodda məsələ həmişə müəyyən bir məqsəd üçün təcrübələr aparıb, ölçmələr edib sonra da müəyyən nəticələrə gəlmək olub. Neyroinformatika metodları nəhayətində bundan fərqli heç nə eləmir; burada edilən əsas şey son bir neçə onillikdə inkişaf etdirdiyimiz müasir metodlarla alınan böyük həcmdə datanın yaratdığı özünəməxsus hesablama problemlərini həll etməkdir. Başqa sözlə, bio/neyroinformatika real dünyadan rəqəmsal dünyaya keçid etmiş informasiyanı **müşahidədən biliyə çevirməyə çalışır**.

"Çox yaxşı, çox əla" deyə bilərsiz, "bəs anlamağa çətinlik çəkdiyiniz qədər datanı toplamağın mənası nədir?". Yəni, böyük həcmdə datanın yığılması bizə nə verir? Gəlin, bu haqda birgə fikirləşək.

{İnteraktiv müzakirə}

Ən bəsitindən, öyrənməyə çalışdığımız şey haqda mümkün qədər çox və ətraflı müşahidələr toplamaq **o şeyi daha dəqiq görməyimizə imkan verir**. Lakin bundan əlavə, çoxsaylı, eynicinsli datanı bir araya gətirmək həm də bu data toplusundakı **mənəli qənanuyğunluqları görməyə** imkan verir. Məhz bu səbəbdən informatika metodlarının tətbiq olunduğu biologiya sahələri **aşağıdan-yuxarı düşüncənin** məhsulu olan çalışmalar

ətrafında cəmləşir. Bəs reallıqda bu necə baş verir? Bu gün bu suala bir tarixi, bir də daha müasir örnəklə cavab verməyə çalışacağam.

2. Charles Darwin-in Böyük Datası

Bir qədər zarafatla da olsa, Charles Darwin-i bu gün "big data" dediyimiz, müəyyən bir sual ətrafında sistematik şəkildə irihəcmli data toplusu yığmağa çalışan çalışmalarının banilərindən biri adlandıra bilərik. Darwin məşhur Beagle gəmisi ilə 5 il çəkəcək (1831-1836) dünya səyahətinə çıxanda cəmi 22 yaş var idi. Darwin bu səyahəti çoxsaylı müşahidələr aparıb çox sayda nümunələr toplamağa sərf etmişdir.

Səyahətin bitmə tarixindən 1859-cu ildə dərc olunan Növlərin Mənşəyinə qədər keçən 23 ildə o, təbii seçmə yolu təkamül nəzəriyyəsinin doğruluğunu təyin etmək üçün dayanmadan data toplamışdır. Düz 23 il! Bu 23-ilin sonunda belə ona təbii seçmə nəzəriyyəsini izah edən kitabı yazmağa vadar edən şey kifayət qədər data toplaması deyil, eyni nəticələrə daha sürətlə gələn Alfred Wallace-dan gələn məktub olmuşdur! Kim bilir, bəlkə də Wallace olmasaydı Darwin ömrünün axırına qədər data toplamağa davam edərdi. :) (Qeyd: Əlbəttə, 1836-cı ildən sonra Darwin təkə data toplamağa davam etməyib, başqa çalışmalarını və hətta Beagle səyahətinin xronikasını da dərc etdiyib, amma Növlərin Mənşəyini onun əsas nəzəri elmi işi olaraq görə bilərik.)

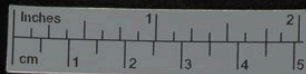
Yaxşı bəs Darwin-in bu yanaşması bu gün bioinformatika dediyimiz sahədəki yanaşmalarla necə üst-üstə düşür?

{İnteraktiv müzakirə}

Bunu başa düşmək üçün ilk öncə Darwin-in Beagle-da səyahət zamanı nə etdiyini başa düşməliyik. Bu günkü bioinformatika çalışmaları kimi, bu hekayənin iki əsas hissəsi var: 1) datanın sistematik şəkildə toplanması və 2) datadakı mövcud olan qanunauyğunluqları nəzərə alaraq uyğun bir metodla onun analizi.

Datanın sistematik olaraq toplanması dedikdə nə nəzərdə tutulur?

Datanın sistematik toplanması deyəndə müəyyən sual üçün **vacib olduğunu düşündüyünüz dəyişənləri** müqayisə oluna biləcək formada toplamaq nəzərdə tutulur. Darwin-in dəftərlərindən onun topladığı data ilə bağlı çox ətraflı detallar əldə edə bilirik. Bu materiallara baxanda görürük ki, canlıların mənşəyi ilə maraqlanan biri olaraq o, müşahidə etdiyi canlıların müxtəlif xassələrini, morfoloji xüsusiyyətlərini, davranışını, necə çoxaldığını, o cümlədən, onun harada yaşadığını, ərazinin iqlimi, və s. sualla əlaqəli olduğu düşünülmə faktörlerle bağlı qeydlər aparmışdır. Məsələn, bu Beagle-da toplanan nümunələrin kataloqlarından birindən götürdüyüm səhifədir:



1832.

Shells in Spirits of wine.

337. *Chondrus*. *Aelix*. *Bulimus*, under stones on Mount. *Chondrus* animal "wood brown" colour. In = *Vides*. August.

412. *Bucanum*. (with ovules) Crawling on the sand banks and living on dead fish; foot oblong, rounded anteriorly, the yellow operculum is placed obliquely on the upper part of the extremity. Siphon lead colour, not closed; tentacula same colour pointed; mouth projecting over foot and between tentacula when closed with small longitudinal divisions; from this there can be protruded a very long, red coloured proboscis, terminal orifice with cartilaginous rim. Very commonly ^{found} on the shoals, there are several ovules, these are about $\frac{1}{16}$ in diameter, rounded conical, with broader base, semi-transparent, on the summit is a circular lid, which falls.

The situation of the ovules or eggs on the shell must be almost necessary, as the animal inhabits extensive sand banks, where there is no hard substance to fix them on.

Bahia. Blanca. Sep^r.

417. *Mya*; dug out of the mud or arenaceous clay bank, 6 inches within; in numbers. B. B.

429. *Crepidula*. Adhering to the anchor, sounding 10 fathoms; shell with curved grooved spines; The young shells adhere to the old one, in these places the spines are absent; animal with foot rounded, posterior half lying, on the diaphragm of shell. Tentacula pointed with minute black eyes situated near the base and on them; between these the mouth opens.

Cambridge University Library

412. Bucanum. (with ovules) Crawling on the sand banks and living on dead fish; foot oblong, rounded anteriorly, the yellow operculum is placed obliquely on the upper part of the extremity. Siphon lead colour, not closed; tentacula same colour pointed; mouth projecting over foot and between tentacula when closed with small longitudinal divisions; from this there can be protruded a very long red coloured proboscis, terminal orifice with cartilaginous rim. Very commonly found on the shoals there are several ovules. these are about 1/12th in diameter, rounded, conical, with broader base, semi-transparent, on the summit is a circular lid, which falls.

The situation of the ovules or eggs on the shell must be almost necessary, as the animal inhabits extensive sand banks, where there is no hard substance to fix them on. Bahia Blanca. Sept

Bunun kimi saysız hesabsız müşahidələrin toplandığı onlarca dəftər olduğunu düşünün. (Qeyd: Maraqlananlar bu əlyazmaların rəqəmsallaşdırılmış nüsxələrini Darwin Online portalından tapa bilərlər.) Bu müşahidələri Darwin indi aparsaydı onları təxminən belə bir cədvələ düzərdi (qeyd: bu tamamilə xəyali bir cədvəldir):

Növ	Ölçüsü (cm)	Rəngi	Qidası	Yaşadığı yer
1	10	göy	yosun	Galapagos
2	3	qırmızı	plankton	Cape Verde
3	4	sarı	plankton	Cape Verde
4	12	mavi	yosun	Galapagos
5	2	narıncı	plankton	Cape Verde

Bu tipli çox sayda növü dərinədən müşahidə edib kataloqladıqdan sonra Darwin-i təbii seçmə yolu ilə təkamül nəzəriyyəsi nəticəsinə gətirən əsas müşahidələrdən biri coğrafi olaraq bir-birinə yaxın yerlərdə yaşayan canlıların mühit baxımından bir-birilə eyni şəraitdə yaşayan canlılardan daha yaxın olmaları faktı idi. Bu isə öz növbəsində onu göstərirdi ki, canlılar mühit şəraiti və yaxud onları yaşadıkları yerə mükəmməl şəkildə uyğun formada Allah tərəfindən deyil, müəyyən tarixi hadisələrin ardıcılığın nəticəsi olaraq formalaşır və dəyişir. Biologiyada tarixi hadisələr dedikdə söhbət canlıların davamlılığını təmin edən irsiyyət əlaqələri, nəsildən nəsilə keçiddən gedir. Əgər mühit şərtləri növ zənginliyini izah edən əsas şey olsaydı, bir-birinə oxşar mühitlərdə (Qalapaqos və Kabo-Verde adalarında) yaşayan növlər bir-birinə daha çox oxşamalı idilər. Lakin bunun əksinə, Darwin müəyyən

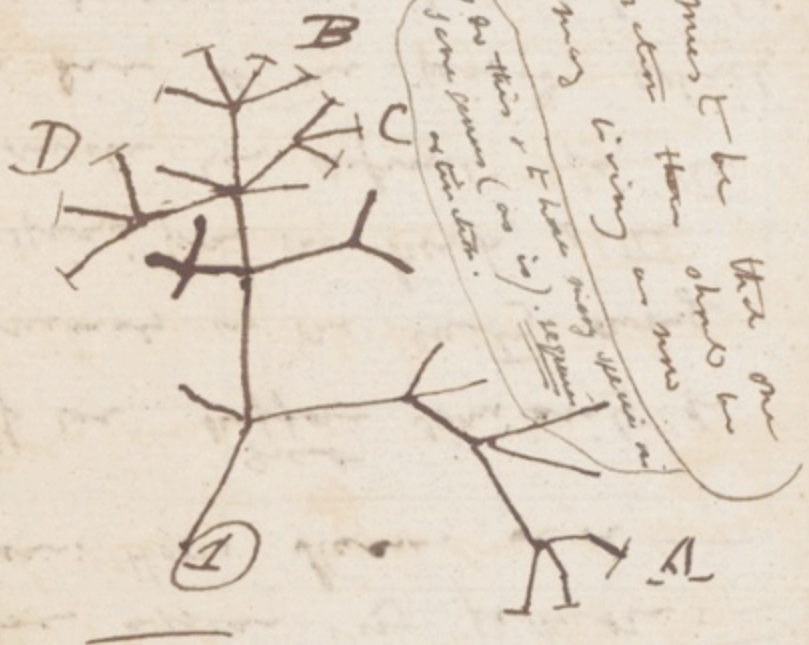
etmişdir ki, Qalapaqosdakı növlərin çoxu ona daha yaxın yerləşən Cənubi Amerika qitəsindəki növlərə daha çox oxşayır (eyni şey Kabo-Verdedə keçərli idi; bu növlər öz növbəsində Afrikadakı növlərlə daha yaxından əlaqəli idilər).

Növ	Ölçüsü (cm)	Rəngi	Qidası	Yaşadığı yer
1	10	göy	yosun	Galapagos
4	12	mavi	yosun	Galapagos
2	3	qırmızı	plankton	Cape Verde
3	4	sarı	plankton	Cape Verde
5	2	narıncı	plankton	Cape Verde

Beləliklə, o bu nəticəyə gəlmişdir ki, Qalapaqos adasında yüksək çeşidlilik nümayiş etdirən növlər əslində ortaq bir əcdada gedib çıxır və o vaxtdan bu yana yaşadıkları mühitin şərtlərinə uyğunlaşaraq fərqliləşiblər.

Növlər arasındakı bu əlaqəni göstərmək üçün o ilk dəfə **dendrogram** dediyimiz və yaxud həyat ağacı olaraq bilinən diaqramı çəkmişdir ki, bu metod bu günədək bioinformatikada eynicinsli şeylər (məsələn, növlər, genlər, proteinlər, hüceyrələr və s.) arasında əlaqəni göstərmək üçün tez-tez istifadə olunur.

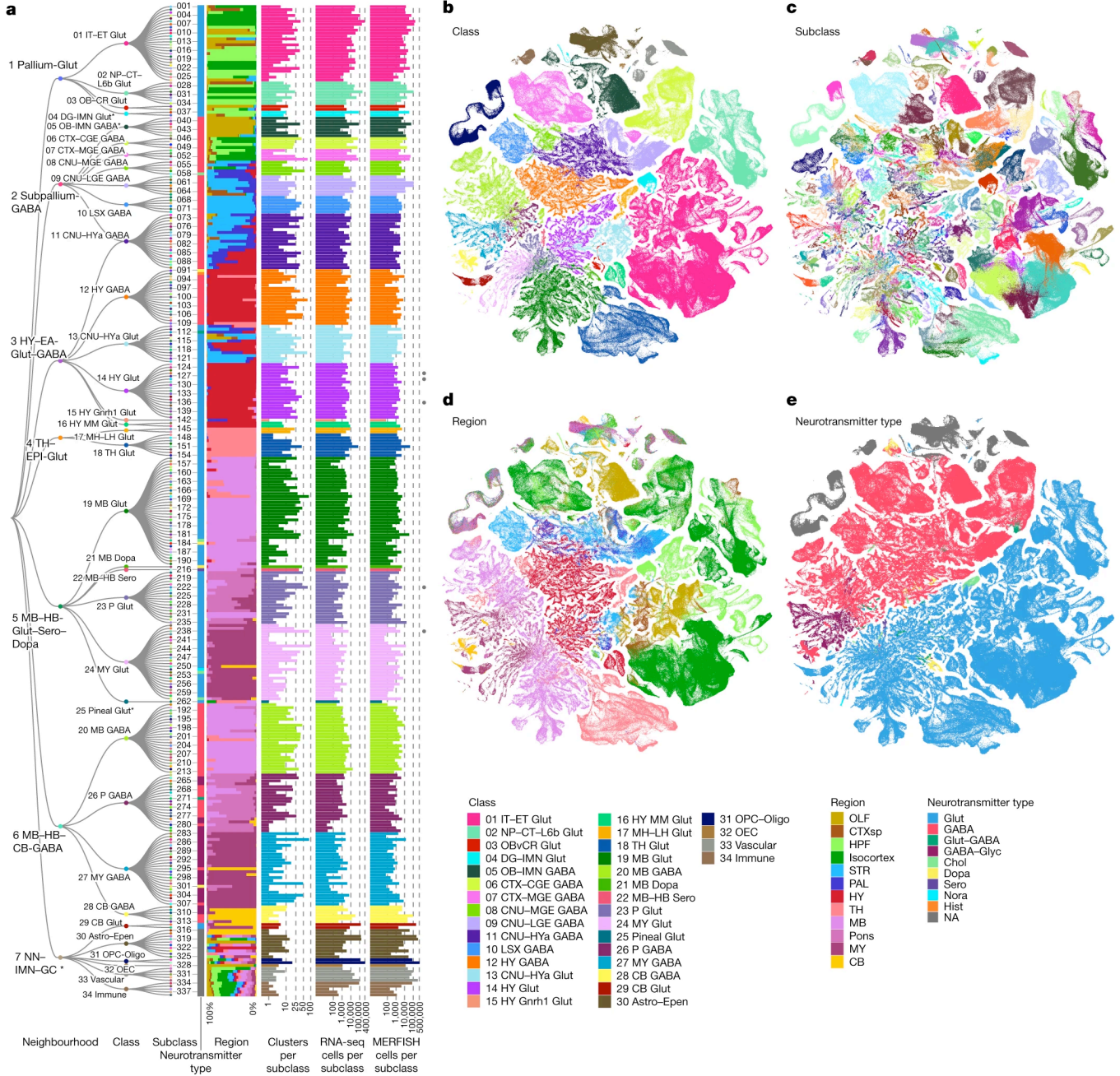
I think



than between A & B. various
 sort of relation. C + B. The
 finest gradation, B & D
 rather greater distinction
 than genera would be
 formed. - bearing relation

to ancient types.

İyerarxik dəstələmə (klasterinq olaraq da bilinir) adlanan bu metod ayrı-ayrı müşahidələrin arasında bənzərliyi ölçüb, ən bənzər olanları bir dəstəyə qoyur və prosesi bütün müşahidələr hər şeyi ehtiva edən ümumi bir dəstənin parçası olana qədər davam etdirir. Məsələn, beyində fərqli hüceyrələrin hansı genləri ifadə etdiklərini ölçdükdən sonra bu metoddan istifadə edərək onlar arasında əlaqələri, hansı ümumi hüceyrə növlərinin olduğunu öyrəne bilərsiniz:



([Yao et al. 2023](#))

Darwin dendroqramı yaratmaq ideyasını beyində, intuitiv yolla etsə də müasir neyroinformatikada dendroqramları yaratmaq üçün alqoritmlər mövcuddur və bunlardan istifadə edərək nəhəng data toplularında eynilə Darwin-in elədiyi kimi mövcud oxşarıqlar və qohumluq əlaqələrini tapmaq mümkündür.

3. Epilepsiyası olan xəstə üçün neyroinformatika

İlk bölmədə neyroinformatikanın təməl motivasiyasını izah etmək üçün Darwin-in çalışmalarını və müşahidələrini anlamaq üçün tətbiq etdiyi məntiqi izah etməyə başladıq. Eyni yanaşmanın biologiyada və neyroelmdə bir çox problemə tətbiq edə bilərik. Bunun hal-hazırda tibbi diaqnostika və müalicədə istifadə olunan real örnəklərini görmək üçün xəyali, epilepsiyası olan bir körpənin misalına nəzər salacağıq. **Bio/neyroinformatika bu xəstədə diaqnoz üçün nə təklif edir?** Bunu anlamaq üçün gəlin xəstənin ilkin qəbulundan sonra müalicə seçiminədək olan mərhələlərdə istifadə olunan (və ya gələcəkdə olunacaq) metodlara nəzər salaq.

3.1 Xəstənin ilkin qəbulu - EEG

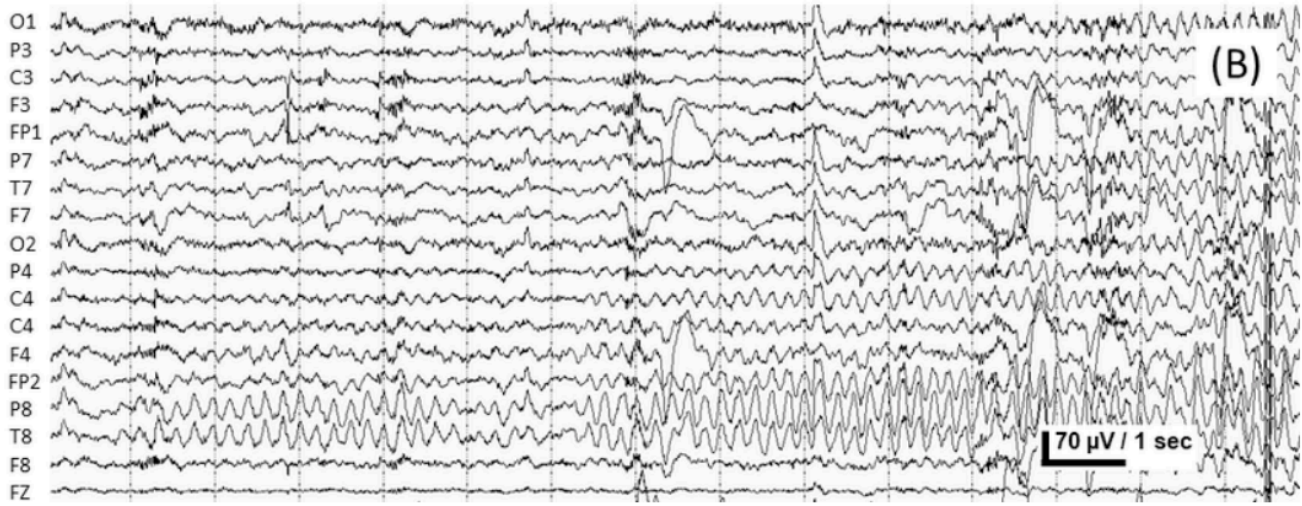
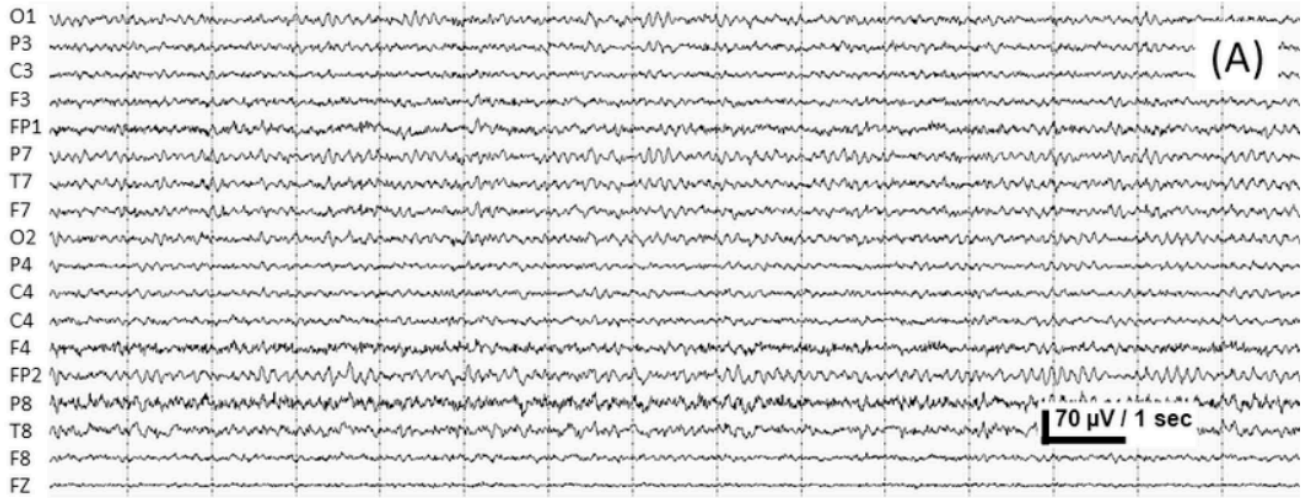
Epileptik tutmaları olduğu düşünülen körpə epileptoloqun qəbuluna gələndə ediləcək ilk işlərdən biri tutmaların həqiqətən də epileptik tutma olub-olmadığını təyin etməkdir. Bunun standard metodu Şəmsin də əvvəlki dərslərdən birində müzakirə etdiyi elektroensefaloqram (EEG) ilədir. Qısaca, EEG-dən istifadə edərək beynin elektrik fəaliyyəti ölçülən zaman kəllə səthinə 32 elektrodlu EEG papağı bərkidilir. Bu elektrodlar saniyədə 256 ədəd ölçmə aparır. Araşdırmalar göstərir ki, uzunmüddətli EEG ölçmələri iktal hadisələri (tutmaların səbəbi olan neyron fəaliyyətlərini) təsbit etmə şansını əhəmiyyətli dərəcədə artırır və diaqnostik səmərəliliyi artırır. Buna görə də klinikalarda iktal hadisələri təsbit etmək üçün standart olaraq 24-saatlıq video-EEG monitorinqi sessiyaları aparılır.

Proses belə görünür:

<https://www.youtube.com/watch?v=p38nzOGJZtl>

Bu yolla günə $32 \text{ elektrod} \times 256 \text{ ölçmə} \times 86,400 \text{ saniyə} = 708 \text{ milyon müşahidə}$ toplanır. Bu monitorinqlər çox vaxt 3-5 gün çəkdiyinə nəzərə alsaq adi bir xəstə üçün ümumilikdə 2.1-3.5 milyard müşahidə toplanır. Xəstədən məsul kliniki neyrofizioloq ənənəvi olaraq bu uzun ölçmə sessiyalarını gözlə incələyərək iktal hadisələri və digər EEG xassələrini təyin edir. Həkimin burada məqsədi bir dənəsi təxminən 90 saniyə (~370,000 müşahidə, yeni ümumi müşahidələrin ~0.0124%-ini) çəkən tutmaları təsbit edib onun xassələrini anlamaqdır. Xüsusilə də tutmaların çox güclü olmadığı və videodan gözlə görülmədiyi hallarda bu tapşırıq saman topasında iynə axtarmaq kimi bir şeydir.

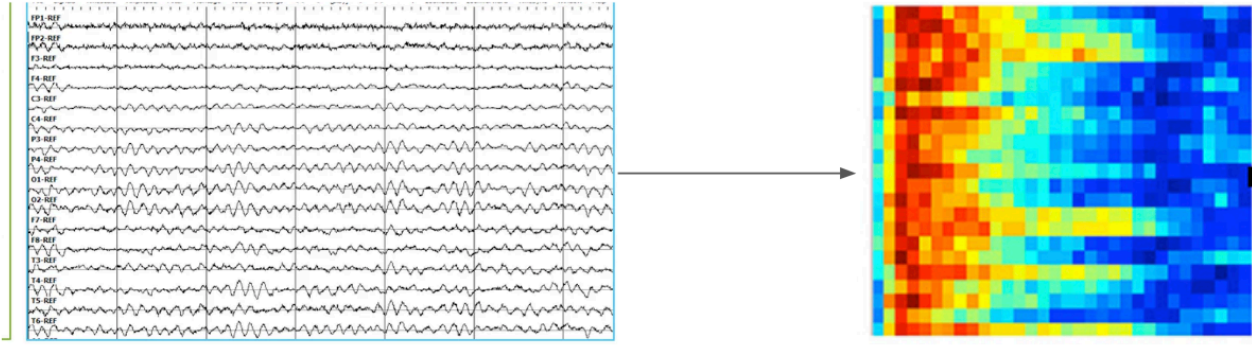
Tapşırığı bir az vizual başa düşmək normal və tutma EEG-sini qarşılaşdırdıqda görə bilərik.



(Fürbass 2017)

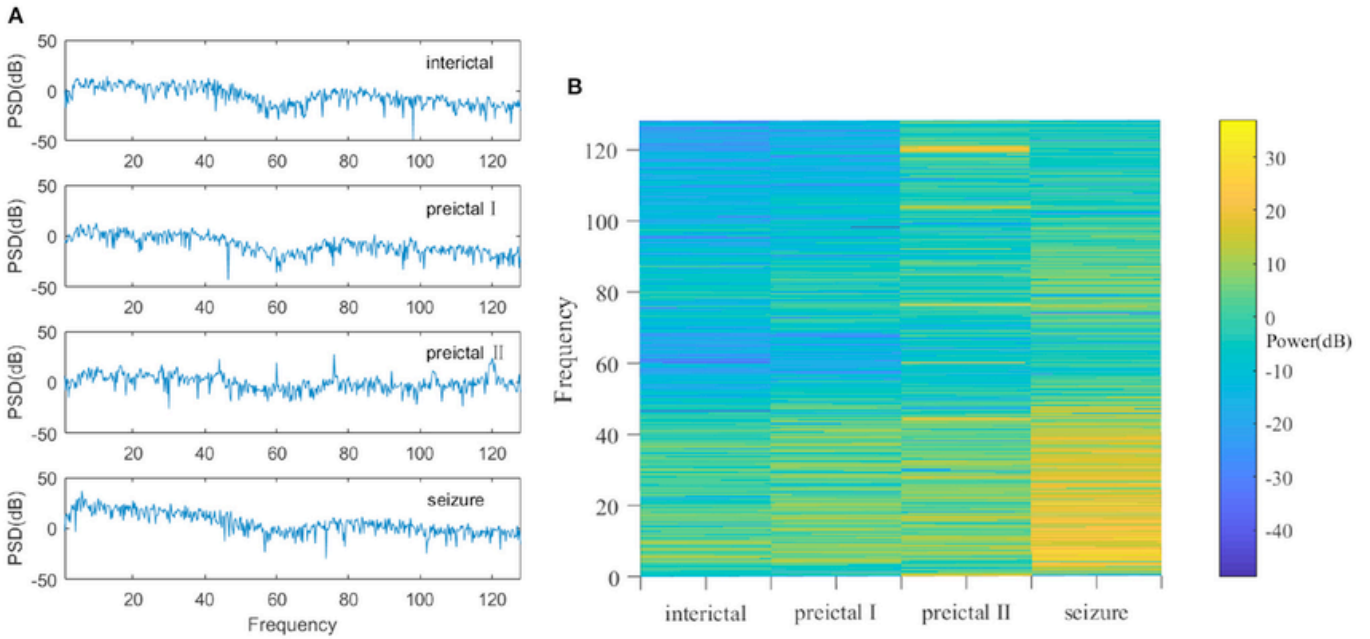
Data miqdarının çox insan vaxtının məhdud olduğu yerlərdə neyroinformatika metodları köməyə gəlir. Bəs bu qədər böyük həcmdə datadan mənalı bir nəticəyə gəlmək üçün bizə nə lazımdır? Ənənəvi olaraq belə problemlərdə maşın öyrənməsi metodlarından istifadə olunur. Bu metodların inkişaf etdirilməsi üçün ən vacib inqredient isə həqiqətdən əmin olduğumuz istinad databazalarıdır.

Artıq uzun illərdir toplanmaqda olan EEG ölçmələri bəzi data bazalarda bir araya gətirilir. Məsələn, bunlardan ən böyüklərindən biri olan [Temple Universiteti EEG Korpusudur](#) ki, burada universitetin hospitalında 2002-2017-ci illər arasında toplanmış 26,846 kliniki EEG yazıları toplanmışdır. Kompüterlər iktal hadisələri təyin etməyi necə öyrənirlər? Bunun üçün ilk növbədə datanı kompüterə "başla düşə biləcəyi" hala gətirmək lazımdır.



(Gao et al. 2020)

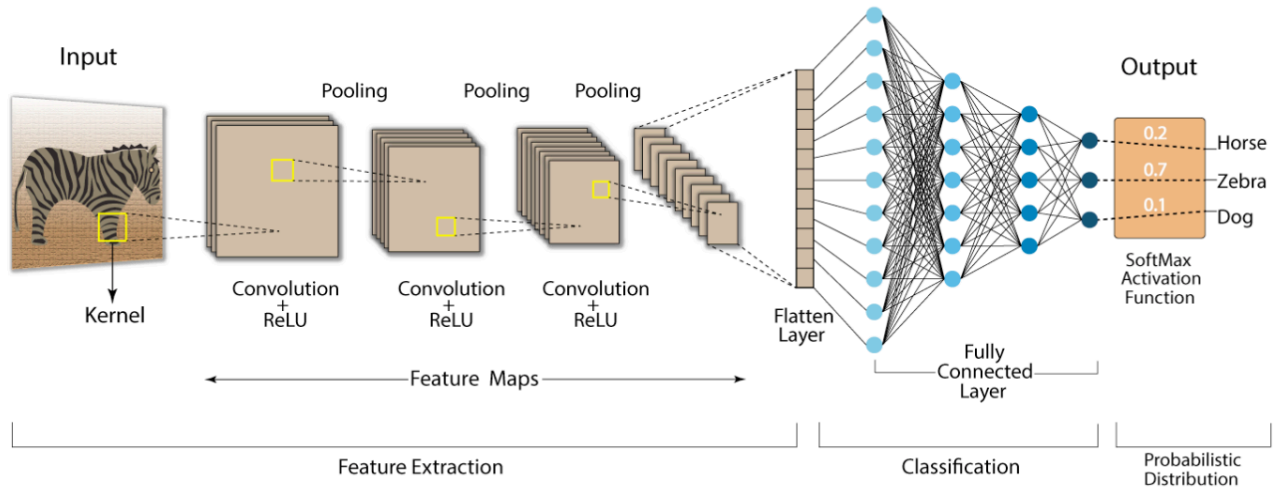
Bu addım nəyə görə lazımdır? Çünki epileptik tutmalar zamanı EEG ilə ölçülən beyin fəaliyyəti müəyyən xarakteristik xüsusiyyətlərə malik olur. Bunu görmək üçün elektrodlardan gələn siqnal fərqli tezliklərdəki fəaliyyət dərəcələrinə ayrılır (buna Fourier transformasiyası deyilir). Siqnala bu şəkildə baxanda tutma zamanı aşağı tezliklərdə fəaliyyətin gücləndiyini görmək mümkündür.



Maşın öyrənməsi metodunun bu qanunauyğunluğu tutması üçün çox vaxt datanı tezliklərə ayırıb o şəkildə analiz edirik. Öyrənmə mərhələsinin özü üçün bir sıra metod olsa da, hazırda bu tapşırıqda istifadə olunan ən qabaqcıl maşın öyrənməsi modelləri konvolyusiyalı neyral şəbəkə (convolutional neural network, CNN) metoduna əsaslanır. CNN-lər hər saniyə alınan 32 elektrod \times 256 ölçməni bir növ "şəkil" kimi qəbul edir.

Alınan şəkil beyindəki görmə korteksindəki hüceyrələrin həyata keçirdiyi əməliyyatları müəyyən mənada təqlid edən süni hüceyrələr tərəfindən analiz olunur. Süni neyronlar CNN-lərdə fərqli qatlar boyu düzülür və model tapşırığı öyrəndikcə, bu neyronlar da datadakı fərqli naxışlara qarşı həssas olmağa başlayır.

Convolution Neural Network (CNN)



Əlbəttə tutmalar insandan-insana, bölgədən-bölgəyə, hətta tutmadan-tutmaya dəyişdiyinə görə standard bir "şəkli" əzbərləmək kifayət etmir. CNN-lər siqnaldakı mürəkkəb, insan gözü ilə seçilə bilməyən əlaqələri öyrənir və hər bir şəklin hansı perioda (tutma, qeyri-tutma, və s.) aid olduğunu sinifləndirir.

Pasiyentdən alınan dataya bu analiz tətbiq olunduqdan sonra CNN həkim üçün onun fokuslana biləcəyi vacib anları seçir və beləliklə böyük vaxt sərfiyyatının qarşısı alınır. Hazırda tutmaların təsbiti üzrə bazarda olan ən yaxşı model Persyst adlanan şirkət tərəfindən istehsal olunan modeldir ki, bu modelin dəqiqliyi insanlarınkına yaxındır. Zamanla bu tapşırıqla bağlı data artdıqca mövcud alqoritmlər tapşırığı yəqin ki insanlardan da daha yaxşı yerinə yetirəcək və bu yolla tutmaların təsbiti və xarakterizasiyası daha da dəqiq hala gələcək.

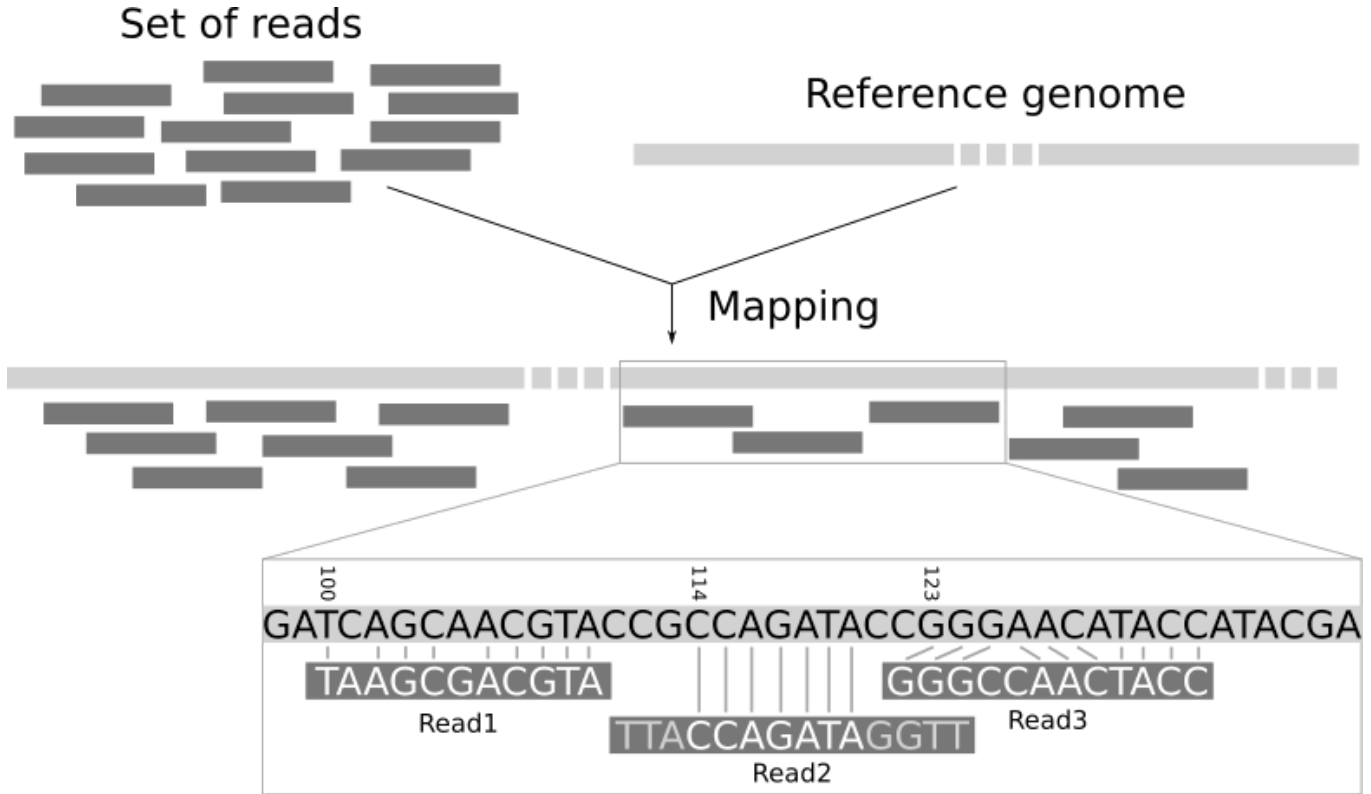
3.2 Genetik analizin sifarişi - Gen ardıcıllanması

Tutmaları təsbit olunan körpədə tutmaların nədən qaynaqlandığını anlamaq üçün edilən ən birinci növbəti addımlardan biri genetik analizlərin aparılmasıdır, çünki erkən epilepsiyalarda genetik etiologiyanın ehtimalı böyük olur. Genetika və buna bağlı olaraq bioinformatikada son iyirmi ildə baş verən inkişaf insan genomunun ardıcılmasını \$1000-dan aşağı salmağı bacarmışdır. Laboratoriya metodu ucuzlaşdığına görə ondan istifadə də çoxalır. İstifadənin çoxalması ilə bu yolla yaranan datanın həcmi də artır. Beləde yenə də bioinformatikanın köməyindən yararlanmaq üçün şərait yaranır.

Gəlin görək, ilk öncə, bu ardıcılalma prosesi tam olaraq necə baş verir. Prosesin ilk mərhələləri ilə bağlı anlayış yaratmaq üçün bu videoya baxa bilirik:

<https://www.youtube.com/watch?v=2JUu1WqidC4>

Prosesin bizə maraqlı olan hissəsi videoda səs dalğaları ilə uzun DNA zəncirlərinin kiçik fraqmentlərə parçalanması və "güclü kompüterlərin bu fraqmentləri birləşdirməsi" deyə üzərindən sürətlə keçilən hissədir. Kompüterlər milyonlarla qısa DNA ardıcılığını 6 milyard nukleotid cütü olan insan genomuna necə çevirir? Müasir növbəti nəsil genom ardıcillamasında orta hesabla genomun hər nöqtəsi ~15-30 dəfə oxunur. Bunun nəzərə alsaq problemin mürəkəbbliyini başa düşməyə başlaya bilərik.



Yeni, qarşımızdakı tapşırıq bu xırda fraqmentlərin hər birinin 6 milyardlıq istinad genomunda hansı nöqtəyə uyğun gəldiyini (başlanğıc və bitmə koordinatlarını) tapmaqdan ibarətdir. Bunu bir növü mətn axtarışı problemi olaraq görə bilərsiniz. Axtardığınız "söz" qısa fraqment, "lüğətiniz" isə bütöv insan genomudur.

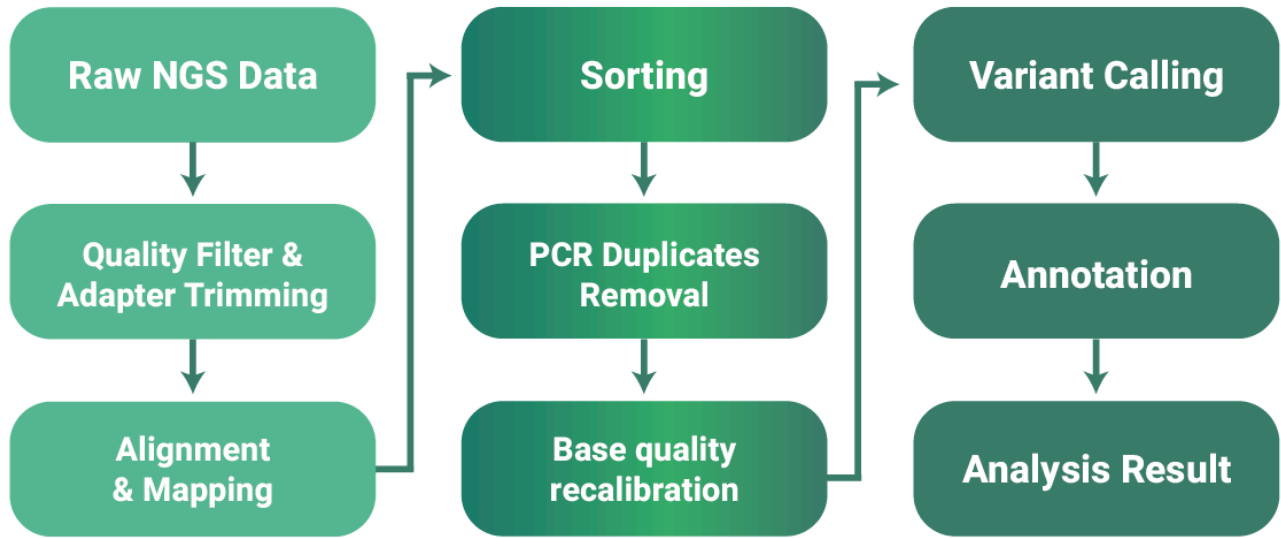
Bu tapşırığı etmək üçün istifadə olunan Burrows-Wheeler Hizalayıcısının (Burrows-Wheeler Aligner, BWA) tam olaraq nə etdiyini izah etmək bir az uzun vaxt alacaq, ona görə də detallarını müzakirə etməyəcəyik. Sadəcə konseptual səviyyədə bu alqoritm hər fraqmentin istinad genomdakı mümkün yerini tapmaq üçün hər fraqment üçün bütün genomu hərfbəherf oxumur. Alqoritmin etdiyi şey ilk öncə axtarış prosesini asanlaşdıracaq bir axtarış indeksi yaratmaqdır. İndeksi bir növü axtarışı sürətləndirən bir siyahı kimi düşünə bilərsiniz. Məsələn, necə ki, kağız lüğətdə "epilepsiya" sözünü axtarmaq üçün bütün lüğəti oxumaq yerinə "E" hərfinin olduğu yerə gedirsiniz, bu yaradılan indeks də əlinizdəki fraqmenti tapmağı rahatlaşdırır.

İndeksin yaradılması Burrows-Wheeler Transformasiyasını ilə edilir ki, bu transformasiyanın etdiyi şey əslində çox asandır:

Transformation				
1. Input	2. All rotations	3. Sort into lexical order	4. Take the last column	5. Output
<div style="border: 1px solid gray; padding: 5px; width: fit-content; margin: 0 auto;"> $\text{^BANANA\\$}$ </div>	$\text{^BANANA\$}$ $\text{\BANANA $\text{A\BANAN $\text{NA\BANA $\text{ANA\BAN $\text{NANA\BA $\text{ANANA\B $\text{BANANA\$^}$	$\text{ANANA\B $\text{ANA\BAN $\text{A\BANAN $\text{BANANA\$^}$ $\text{NANA\BA $\text{NA\BANA $\text{^BANANA\$}$ $\text{\BANANA	$\text{ANANA\B $\text{ANA\BAN $\text{A\BANAN $\text{BANANA\$^}$ $\text{NANA\BA $\text{NA\BANA $\text{^BANANA\$}$ $\text{\BANANA	<div style="border: 1px solid gray; padding: 5px; width: fit-content; margin: 0 auto;"> $\text{BNN^AA\\$A}$ </div>

Bəs bu indeks "ANA" və ya "BAN" tipli fraqmentləri sürətlə tapmaq üçün necə istifadə olunur? Bu suala cavabı ev tapşırığı olaraq vermək istəyirəm. Metodun dərc edildiyi [orijinal məqalə](#) və ya YouTube videolarından (1 - transformasiya üçün <https://www.youtube.com/watch?v=Lc-ACiJlrnM>, 2 - axtarış üçün <https://www.youtube.com/watch?v=uKreghMwLLE>) baxıb bu metodu öyrənməyi təklif edirəm, çünki prosesi anlamağın verdiyi həzz başqa şeydir :)

DNA fraqmentlərinin istinad genoma hizalanması ən birinci mərhələlərdən biridir. Bu mərhələdən sonra bir neçə mərhələ var ki, burada istifadə olan metodların hər biri özlüyündə yüzlərlə məqalənin mövzudur. Körpənin genom ardıcılığı təyin olduqdan sonra əlimizdə olan fərdə aid nukleotid ardıcılığı istinad genom ardıcılığı ilə qarşılaşdırılır və fərdə xas genetik variantlar (yeni istinaddan kənar çıxımlar) təsbit olunur. Bioinformatik metodların inkişafı əvvəllər günlər alan bu analizi standartlaşdırmış və bir neçə saata endirmişdir. Hal-hazırda mövcud metodlar yüksək dəqiqliklə genomdakı variantları tapa bilir.



Bu analizin nəticəsində əlimizdə adam başına təxminən 4-5 milyon variant qalır. Fikir verdinizsə, 6 milyard nukleotiddən 4-5 milyona endik, amma bu hələ də insanların analiz edə biləcəyindən çox daha böyük bir rəqəmdir. Bunu necə aşağı sala bilərik?

Genomun böyük hissəsi (98%+) birbaşa olaraq proteinləri kodlaşdırmadığına görə bu variantların böyük hissəsi hüceyrədəki proteinlərin quruluşuna təsir etmir. 4-5 milyon variantın arasından proteinlərə birbaşa təsir edə biləcək mutasiyaların sayı təxminən 20,000-25,000-dir. Epilepsiya ilə əlaqəli olan genlər sayı (~500) ümumi genlərin sayının kiçik bir hissəsi olduğuna görə bu variantlarımızın sayını daha da azaldıb 500-1,000 varianta salır. Bu nöqtədə xəstəliyə səbəb ola biləcək varianta yaxınlaşsaq da, artıq ədəbiyyata müraciət etmək məcburiyyətindəyik və bu kifayət qədər çox iş deməkdir (kliniki genetik hər bir pasiyent üçün 500 dənə variantı analiz etsə gerek bütün ili o pasiyentin üstündə işləsin).

Namizədləri daha da azaldaraq bir neçə varianta endirmək üçün bir sıra bioinformatika metodlarına müraciət olunur. Bu metodlar daha əvvəl müşahidə olunmamış bir variantın xəstəlik törədib-törətməmə ehtimalını hesablayırlar. Ehtimallardan istifadə edən kliniki genetik əlindəki variantların hansının xəstəliyi törədə biləcəyi ilə bağlı nəticələrə gəlir.

Patogenliyi (yeni xəstəliyə səbəb olma qabiliyyətini) proqnozlaşdıran metodlardan ən qabaqcılı bu gün Google-un DeepMind qrupu tərəfindən 2023-cü ildə inkişaf etdirilən AlphaMissense modelidir. Bu modelin uğurlarının təməlinə DeepMind-in qurucularından olan və bu il Nobel mükafatı alan Demis Hassabis-in rəhbərliyi altında inkişaf etdirilmiş proteinin üçölçülü quruluşunu təxmin edən AlphaFold modeli durur. (Modelin detallarına girməyə vaxt olmasa da, maraqlananlar orijinal məqaləyə baxa bilərlər: [Cheng və b. 2023](#))

Input

Reference:

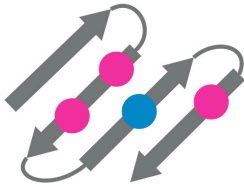
CAG----- DNA
MDVVAMVNR^QTVATMIS----- Protein

Missense variant:

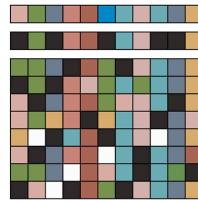
CGG----- DNA
MDVVAMVNR^RTVATMIS----- Protein

AlphaMissense

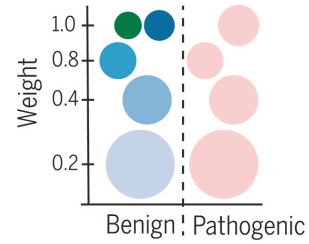
1 Structure context



2 Protein language modeling

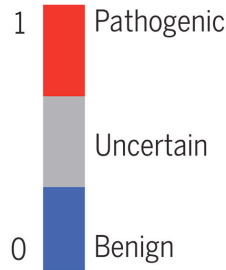


3 Training variants

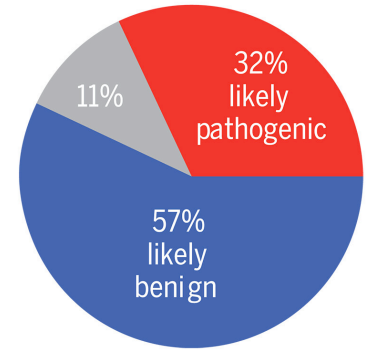


Output

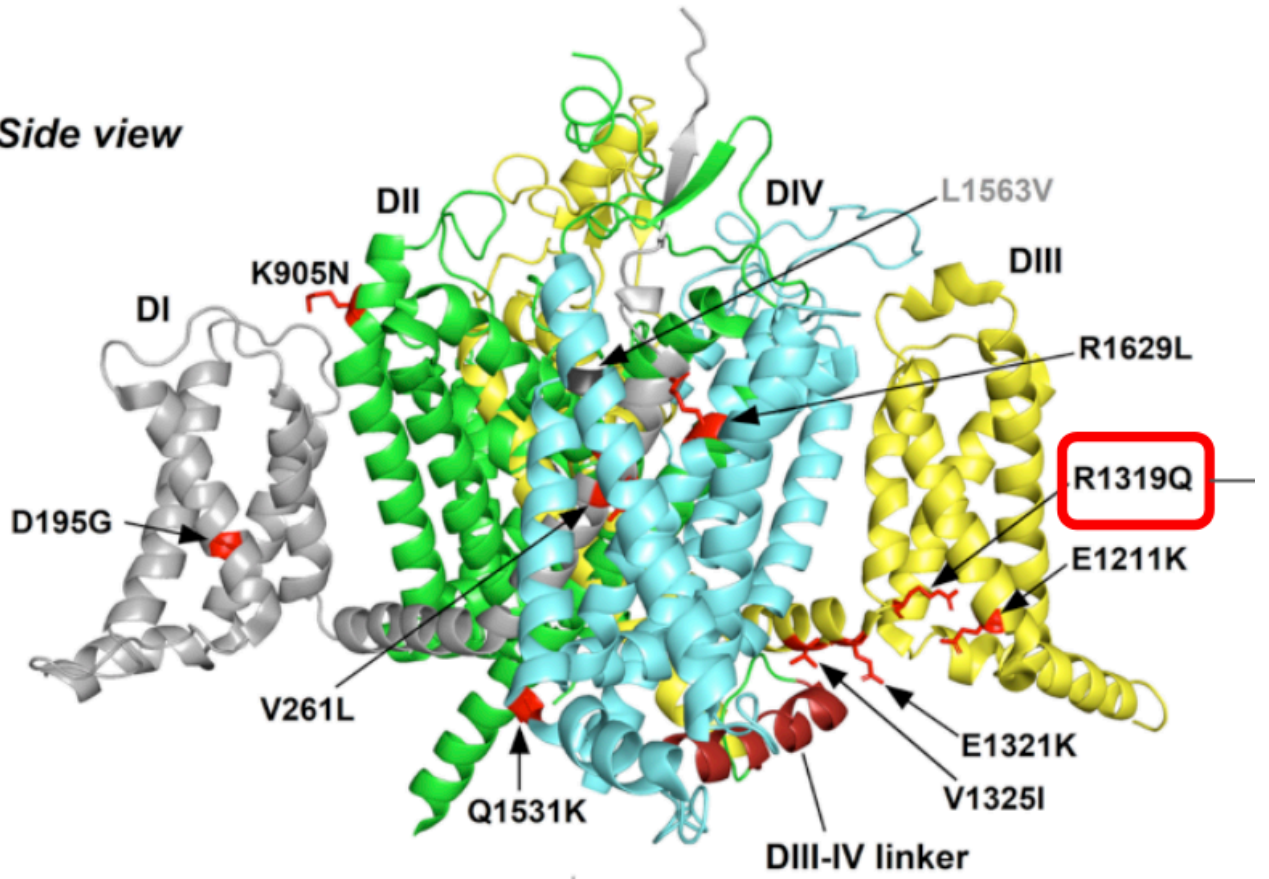
AlphaMissense pathogenicity:



For all 71M possible missense variants in the human proteome:



Ədəbiyyatdan və kompüter modellərindən istifadə edən kliniki genetiklər körpədə müşahidə olunan variantların sayını 1-2 mümkün varianta endirir və onların xəstəliyə səbəb olub-olmadıqları ilə bağlı mülahizə yürüdürlər. Bizim örnəyimizdə fərz edin ki, körpədə SCN2A adlı natrium kanalı kodlayan bir gendəki mutasiyanın xəstəliyə səbəb olduğu qənaətinə gəlirsiniz. Daha dəqiq desək, 1319-cu aminturşu olan arjinin (R) dəyişərək qlutaminə (Q) çevrilmişdir (R1319Q). Əla, bəs bu informasiya ilə nə etməli? Bu bilgi xəstəyə uyğun müalicəni təyin etməyə necə kömək edəcək?

b*Side view*

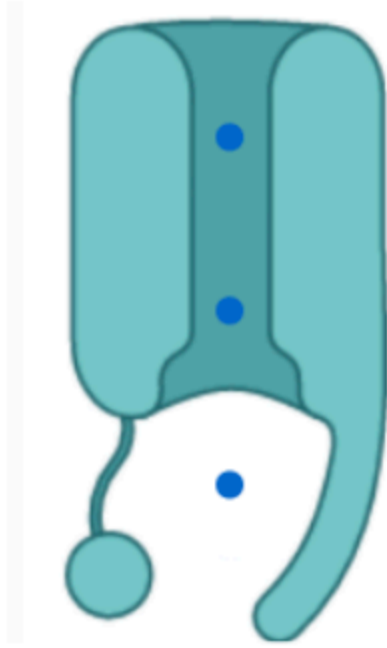
{İnteraktiv müzakirə}

3.3 Funksiyanın itirilməsi yoxsa qazanılması?

Genetik analizin sonunda gəlinən nəticələrdən istifadə etmək üçün ideal olaraq variantın funksional olaraq hansı dəyişikliyə yol açdığını anlamaq istəyirik: *variant hansı yolla xəstəliyə yol açır?*

İon kanallarındaki funksional dəyişmələri əsas iki qrupa ayırmaq olar: funksiyanın qazanılması (gain of function, GOF) və itirilməsi (loss of function, LOF). Qazanma mutasiyalarında ion kanallarından normadan artıq, itirilmə mutasiyalarında isə normadan az ion keçir.

GOF

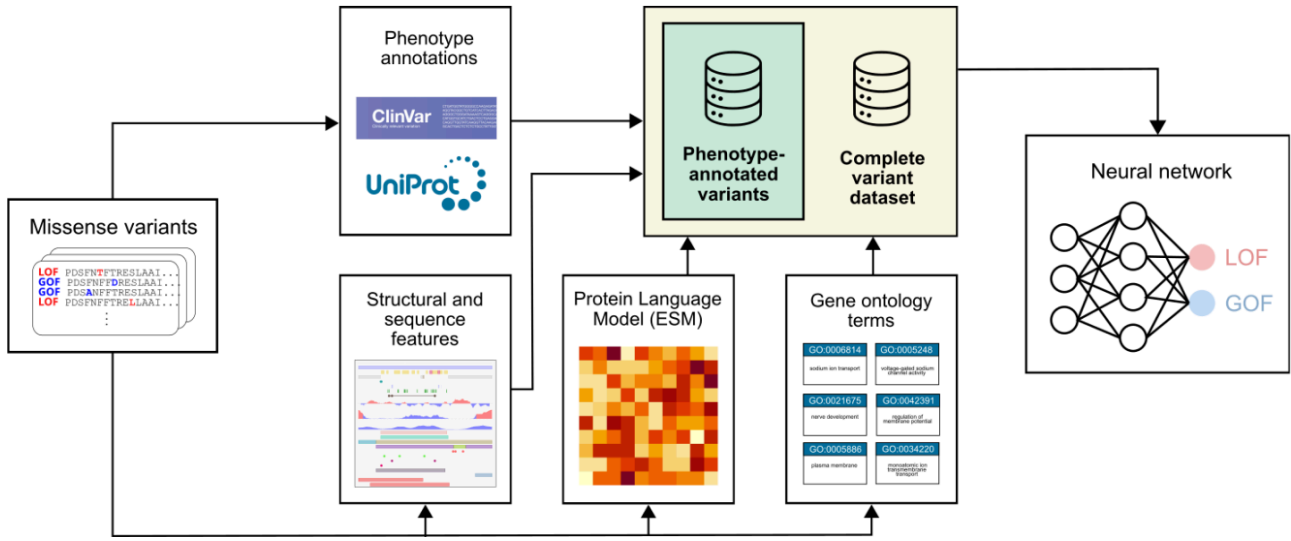


LOF



Kanalın necə dəyişdiyini isə patch-clamp adlanan fizioloji metoddan istifadə edirik. Bunun üçün ilk öncə, kanalın müvafiq variantını kulturada böyütdüyümüz bir model sisteme genetik yollarla köçürürük (ion kanalları üçün bu adətən HEK293 və ya Çin siçancığının yumurtalıq hüceyrələri olur). Daha sonra bu sisteme mikrometr miqyasında kiçik bir mikropipet endirərək hüceyrədə istehsal olunan kanallardan elektrik cərəyanının necə keçdiyini ölçürük.

Ədəbiyyatda çox sayda bu tipli çalışmalar olduğuna görə, EEQ misalında olduğu kimi burada da həqiqətdən əmin olduğumuz məlumatları bir araya topladıqdan sonra maşın öyrənməsi modeli ilə mutasiyaların funksional təsirlərini sinifləndirə bilən bir model inkişaf etdirə bilərik.



Bizim Synaptica-da inkişaf etdirdiyimiz bu model daha önce bəhs etdiyimiz AlphaMissense kimi ion kanallarının üçölçülü strukturu, onların funksional xassələri və digər informasiya növlərindən istifadə edərək ion kanallarındaki aminturşu dəyişmələrinin hansı funksional dəyişməyə yol açdığını (LOF və ya GOF) proqnozlaşdırır. Bu proqnoz müalicədə nə işə yaraya bilər? Əlimizdəki variantın ion kanalları keçiriciliyini artırıb-azaltdığını bildikdən sonra (prinsip etibarilə) ion kanallarına bunun əksinə təsir edən dərmanlardan verə bilərik. Məsələn, mutasiya mutasiya oyandırıcı (excitatory) hüceyrə növlərində natrium keçidini artıraraq elektrik fəaliyyətinin artmasına və bu yolla epileptik tutmalara yol açıbsa, müvafiq kanaldan ion keçidini azaldan dərmanlardan (məsələn, # [lamotrigine](#)) istifadə edərək xəstəliyi birbaşa qaynağından müalicə etməyə çalışa bilərik.

Hal-hazırda bu son mərhələ praktikada olmasa da, yaxın gələcəkdə kliniki genetikanın bu istiqamətdə gedərək genetik variantların daha funksional şərhə ilə praktikaya töhfə verəcəklərini təxmin etmək mümkündür. Şübhəsiz ki, bu yolda əsas rolunu neyroinformatikada inkişaf etdirilən modellər oynayacaq.

Xülasə

Beləliklə bu dərstdə EEQ-də tutmaların təsbiti və DNA ardıcılığının analizində istifadə olunan bir sıra metod üzərindən bio/neyroinformatika metodlarının kliniki praktikada necə real olaraq tətbiq olunduğunu gördük. Burada sadaladığımız şeylər bu gün artıq istifadə olunan şeylər olsa da, bu sahələrdə inkişaf etdirilən metodların çoxu hələ ki, klinik praktikada istifadə olunacaq yetkinliyə çatmayıb. Neyroelmdə elmi səviyyədə gəlinmiş nəticələrin ümumilikdə nəyə görə gözlənilən kliniki təsire malik olmadığı haqda düşünməkdə fayda var, əlbəttə ki. Bunun üçün Nicole Rust-ın bu il dərc olunmuş [Elusive Cures](#) adlı kitabına baxmağı tövsiyyə edirəm. Sahədə çalışanlar üçün bu nə qədər ruhdansalıcı bir nəticə olsa da, yeni daxil olanlar üçün bir o qədər həvəsverici bir şey

olmalıdır, çünkü bu hal neyroelmin kliniki praktikaya sızması üçün imkanların çoxunun gələcəkdə yaranacağına işarə edir.

Əlavə resurslar

- International Neuroinformatics Coordinating Facility. *Introduction to Neuroinformatics*:
<https://training.incf.org/course/incf-short-course-introduction-neuroinformatics>
- <https://www.sciencedirect.com/science/article/pii/S1059131121001989>
-