

PTSDDB:

*Compilació de metadates de biomarcadors
putatius relacionats amb el PTSD*

Josep Marin-Llaó

Tutora acadèmica: Marta Sales-Pardo

Tutor extern: Daniel Domingo-Fernandez

Treball de final de grau de Biotecnologia



UNIVERSITAT
ROVIRA I VIRGILI

Tarragona, 17 de Juny del 2019

Índex

Resum	3
Dades del centre associat al projecte	4
1. Introducció	5
2. Estat de l'art	6
2.1. Transtorn d'estrès posttraumàtic (PTSD)	6
2.1.1. Síntomes psicossomàtics	6
2.1.2. Etiologia	7
2.1.3. Comorbiditat	8
2.1.4. Tractament	8
2.2. Biomarcadors putatius del PTDS	8
2.2.1. Biomarcadors com a indicadors clínics	9
2.2.2. Biomarcadors putatius del PTSD en la literatura	10
3. Objectius	13
4. Materials i Metodologia	13
4.1. Procediment de curació	13
4.2. Model de la base de dades	15
4.3. Desenvolupament de l'aplicatiu web	19
4.3.1. Back-end, modelatge i gestió de la BD: Django i Python	19
4.3.2. Front-end, interfícies: Javascript i OpenAPI	20
4.4. Mesura del consens: Entropia de Shannon	20
5. Resultats	21
5.1. Aplicatiu web PTSDDDB i descriptiva de les dades	21
5.1.1. Resum de freqüències de biomarcadors	22
5.1.2. Direcció del canvi relatiu per biomarcador	23
5.1.3. Anàlisi de la literatura	24
5.1.4. Metadades dels estudis	27
5.1.5. Accés al contingut de la base de dades: API RestFUL	29
5.2. Anàlisis de dades de PTSDDDB	30
5.2.1. Consultes dirigides mitjançant l'API de PTSDDDB	30
5.2.2. Substrats dels biomarcadors putatius més freqüents	30
5.2.3. Anàlisis del consens en el canvi relatiu dels biomarcadors	31
5.2.3.1. Càlcul de l'entropia de Shannon	31
5.2.3.2. Distribució del canvi relatiu dels biomarcadors amb l'entropia	32

Conclusions	34
Reflexió d'autoavaluació i agraïments	35
Llistat de referències	36
Annex 1	38
Annex 2	39
Annex 3	40
Annex 4	41

Resum

S'aborda el repte d'integrar i consensuar la descoberta de coneixement entorn el trastorn d'estrès posttraumàtic, recopilant les evidències experimentals recollides a la literatura, amb un enfoc particular en els biomarcadors putatius de l'enfermetat. Amb aquesta finalitat, s'ha modelat i populat una base de dades integrativa que organitza el contingut de les evidències i les metadades i s'ha desenvolupat un aplicatiu web tot oferint a l'investigador una eina usable i accessible. Finalment, s'estableix un anàlisi del consens entorn la direcció del canvi relatiu dels biomarcadors sota l'efecte de la patologia.

Paraules clau

Trastorn d'estrès post-traumàtic, PTSD, psicopatologia, biomarcador, cortisol, BDNF, base de dades biològica, metadades, estudis clínics, estudi de la literatura, recopilació d'evidències, curació, consens, entropia de Shannon, Django, direcció del canvi relatiu.

Dades del centre associat al projecte

Aquest projecte s'ha desenvolupat en el context de les pràctiques externes realitzades al departament de Bioinformàtica, liderat per *Martin Hoffmann-Apitius*, de l'Institut Fraunhofer per a Algorismes i Informàtica Científica (SCAI). Fraunhofer SCAI forma part del Centre de l'Institut Fraunhofer Schloß Birlinghoven (IZB), situat al Schloss Birlinghoven a Sankt Augustin a l'àrea de Bonn-Colònia.

Concretament, el projecte està vinculat a la col·laboració amb *CohenVeterans*, en la línia de recerca de l'estudi del PTSD, l'article del qual ha estat publicat i estic mencionat com a autor secundari. La tutorització d'aquest treball ha estat duta a terme per l'estudiant de doctorat *Daniel Domingo-Fernández*.

Concretament les tasques que vaig dur a terme en el projecte van ser les de modelatge i població de la base de dades i desenvolupament de la interfície web, així com l'anàlisi de dades posterior proposat per aquesta memòria en particular (*Apartats 4.2, 4.3 i 4.4*), recollint, ordenant i analitzant la tasca realitzada en la fase de curació (*Apartat 4.1*).

1. Introducció

Els avenços realitzats en diferents branques de recerca biocientífica i biomèdica, acompanyats de l'aparició de procediments i tècniques experimentals que generen un volum creixent de dades biològiques, deixen patent la necessitat de consensuar aquest coneixement. El desenvolupament de diferents eines de software i computació dins el marc de la disciplina de la bioinformàtica, permeten l'organització, emmagatzemament, anàlisi i interpretació d'aquestes.

En el projecte del qual se'n deriva aquesta memòria, s'aborda el repte d'integrar i consensuar la descoberta de coneixement entorn el trastorn d'estrès posttraumàtic PTSD (per les sigles en anglès *Post Traumatic Stress Disorder*), una psicopatologia comú que es manifesta en algunes persones després d'experimentar un esdeveniment traumàtic, recopilant les evidències experimentals recollides a la literatura.

D'aquesta manera, s'ha desenvolupat una base de dades fent especial èmfasi en marcadors fisiològics que s'estan estudiant com a biomarcadors putatius en la literatura actual sobre el PTSD, complint metadata pròpia d'estudis clínics i procediments experimentals, entre d'altres. La base de dades conté actualment informació rellevant en 109 articles originals publicats entre 1997 i 2017, i amb un còmput de més de 900 biomarcadors, mesurats d'entre 11 fluids biològics, 9 substrats biològics i 23 indicacions. A més, el contingut emmagatzemat és accessible a través d'una aplicació web tot oferint una mostra descriptiva de les dades amb múltiples visualitzacions interactives. La plataforma ha rebut el nom de PTSDDDB. [1]

S'ofereix, per tant, una visió global dels marcadors fisiològics suportats per metadata clínica, amb l'objectiu de proporcionar als investigadors la possibilitat d'explorar i comparar resultats ràpidament, visualitzant el nivell d'evidència de biomarcadors individuals, així facilitant l'anàlisi i la investigació futura en aquest camp.

Aquesta memòria s'estructura primer introduint de forma resumida en l'estat de l'art les consideracions més consensuades entorn la psicopatologia des d'un punt de vista clínic, la caracterització des d'un enfoc bioquímic i fisiològic dels biomarcadors més estudiats associats al PTSD, i finalment detallant les bases de dades biològiques; a partir d'aquest anàlisi. Aquesta aproximació teòrica permet primer plantejar una sèrie d'objectius a complir en aquesta memòria i també justificar les decisions realitzades alhora d'implementar la base de dades. La tasca de desenvolupament s'exposa en l'apartat de metodologia tot revisant les eines emprades per la seva confecció. Finalment, en l'apartat de resultats es mostren i detallen les funcionalitats oferides en l'última versió de l'aplicatiu web i es planteja un enfocament d'anàlisi del consens entorn les dades recollides a la base dades.

2. Estat de l'art

Les línies de recerca al voltant del trastorn tenen com a objecte d'estudi determinar els processos biològics involucrats en la psicopatologia i/o amb la simptomatologia, així com reexaminar els factors etiològics. Derivat d'aquestes investigacions, es pretenen establir marcadors putatius de la patologia, així com proposar possibles agents terapèutics, tot determinant l'eficàcia d'aquests.

2.1. Trastorn d'estrès posttraumàtic (PTSD)

El trastorn d'estrès posttraumàtic és considerat una malaltia de salut mental, desencadenada per una experiència, un fet o una vivència aterridora que s'ha experimentat o s'ha presenciat.

La majoria de les persones que experimenten situacions traumàtiques poden desenvolupar dificultats transitòries per continuar amb la seva vida quotidiana. No obstant, si els símptomes tendeixen a empitjorar, s'allarguen al llarg de mesos i fins i tot anys i interfereixen en la capacitat de realitzar les tasques diàries habituals, pot derivar en el diagnòstic psiquiàtric de PTSD. [2, 3]

A continuació, es revisen els consensos entorn la patologia des d'un enfoc psiquiàtric i/o psicològic, el qual permet justificar les consideracions fetes alhora modelitzar la base de dades, incloent-hi en aquesta diferents anotacions de rellevància per l'estudi del PTSD.

2.1.1. Síntomes psicossomàtics

La simptomatologia del PTSD pot iniciar-se durant el primer mes després del fet traumàtic, però de vegades pot sorgir passat un any. Els símptomes ocasionen importants problemes en l'àmbit social, laboral o relacional, així com en el funcionament diari habitual; poden comprendre reviviscències, malsons i ansietat greu, així com pensaments incontrolats sobre el fet traumàtic. Aquests poden variar en intensitat amb el temps i etapes de tractament.

Seguint els criteris de diagnòstic del DSM-5 [4], s'ha instaurat una classificació psicossomàtica en quatre grups en base als següents trets: records intrusius, evasió, canvis negatius en el pensament i en l'estat d'ànim i en canvis en les relacions emocional i físiques. A continuació es detallen cadascun d'aquests.

Grup de diagnòstic	Descripció psicossomàtica
Records intrusius	- Records recurrents, involuntaris i ansiosos del fet traumàtic. - Reviure el fet traumàtic com si tornés a succeir (retrospectiva). - Ansietat emocional intensa o reaccions físiques a les coses (objectes, moments, olors...) que recorden el succés traumàtic.

Evasió	- Tractar d'evitar pensar o parlar sobre el fet traumàtic. - Evitar llocs, activitats o persones que recorden el fet traumàtic.
Canvis negatius en el pensament i en l'estat d'ànim	- Pensaments negatius sobre un mateix, d'altres persones o del món. - Pessimisme envers el futur. - Problemes de memòria. - Dificultat en mantenir relacions properes. - Sentir-se allunyat de ls familiars i amics. - Falta d'interès en les activitats que abans li agradaven. - Dificultat per sentir emocions positives. - Sentir-se emocionalment insensible.
Canvis en les reaccions físiques i emocionals	- Sorprendre's o atemorir-se fàcilment. - Estar sempre alerta al perill. - Conductes autodestructives. - Transtorns del somni. - Dificultat de concentració. - Irritabilitat, arravataments d'ira o conducta agressiva. - Sentiments exagerats de culpa o vergonya - Somnis aterradors, amb la temàtica del trauma.

2.1.2. Etiologia

Tot i haver-hi un ampli consens en què la causa del desenvolupament de la patologia és d'origen ambiental, essent l'esdeveniment traumàtic el fet precipitant, cada cop hi ha més evidències que factors de risc biològics i psicosocials poden ser predictors de la instauració dels símptomes, de la seva gravetat i de la seva cronicitat.

L'edat, no obstant, no es considera un factor rellevant en l'aparició de l'enfermetat, ja que pot afectar persones de totes les edats.

Els següents factors es consideren possibles desencadenants d'aquest transtorn:

- Viure un trauma intens o durador.
- Haver sofert trauma en els primers anys de vida, maltractament o abús.
- Tenir un treball que incrementi el risc d'exposició a episodis traumàtics, personal militar o de servei de primers auxilis.
- Tenir altres problemes de salut mental, com depressió o ansietat.
- Tenir problemes de consum inadequat de substàncies com alcohol o drogues.
- Manca d'un bon sistema de suport familiar i amistats.
- Tenir familiars consanguinis amb problemes de salut mental.

Els esdeveniments més freqüents que desencadenen el transtorn són:

- Exposició al combat

- Maltractament físic durante la infantesa
- Violència sexual
- Agressió física
- Sofrir una amenaça amb una arma
- Un accident

Altres successos com incendis, desastres naturals, assalts, robatoris, accidents aeris, tortures, segrestos, diagnòstics mèdics de risc, atacs terroristes i altres esdeveniments extrems.

2.1.3. Comorbiditat

La naturalesa multifactorial del trastorn i la pròpia heterogeneïtat clínica de les psicopatologies, estableix diversos perfils de comorbiditat. D'aquesta manera el PTSD pot ser causa, conseqüència o augmentar el risc a patir altres problemes de salut mental, com els següents:

- Depressió i ansietat.
- Problemes amb el consum de drogues i alcohol.
- Trastorns de l'alimentació.
- Pensaments i accions suïcides.

2.1.4. Tractament

El tractament que s'aplica per a superar el PTSD es basa principalment en psicoteràpia, medicació i/o una combinació dels dos. [3, 5]

Psicoteràpia:

- **Teràpia cognitiva:** Ajuda a reconèixer els patrons cognitius, les formes de pensament. S'utilitza sovint conjuntament amb la teràpia d'exposició.
- **Teràpia d'exposició:** Teràpia conductual que ajuda a fer front a situacions i records de forma eficaç. Resulta útil per a vivències retrospectives i malsons.

Fàrmacs:

- **Antidepressius:** Ajuden a mitigar els símptomes de depressió i ansietat, trastorn del son i la concentració. Els fàrmacs inhibidors de la recaptació de serotonina (ISRS), la sertralina (Zoloft) y la paroxetina (Paxil) s'administren per al tractament del PTDS.
- **Ansiolítics.** Alleugen els símptomes d'ansietat i els problemes relacionats.

2.2. Biomarcadors putatius del PTDS

La fisiopatologia del PTSD afecta múltiples sistemes biològics interconnectats que regulen les funcions de salut física i mental, estant associades amb la comorbiditat. Una extensa quantitat d'investigacions relacionades amb el PTSD han explorat l'ús dels marcadors fisiològics com a biomarcadors discrets vinculats amb aquest trastorn; no obstant això, no s'han identificat fins a la data biomarcadors suposats de PTSD sobre la base del procés d'aprovació normativa per qualificar i validar biomarcadors al voltant de contextos d'ús clínic específics. [6]

2.2.1. Biomarcadors com a indicadors clínics

Els biomarcadors són indicadors biològics que poden mesurar-se i relacionar la seva presència i intensitat amb el desenvolupament d'una malaltia, característiques o canvis fisiològics mesurables i avaluables a nivell molecular, bioquímic o cel·lular que actua com a indicador d'un procés biològic normal o patològic o com a resposta a una intervenció terapèutica. Els biomarcadors mesuren una interacció entre un sistema biològic i un agent de tipus químic, físic o biològic, que és avaluada com a resposta funcional o fisiològica, que passa a nivell cel·lular o molecular i a més a més està associada a la probabilitat del desenvolupament d'una malaltia.[7]

La interacció depèn de les característiques heretades i adquirides de l'individu (o del sistema biològic), i de les circumstàncies de l'exposició.

S'estableix una classificació en tres tipus: [8]

- **Biomarcador d'exposició**, el qual evalua en un organisme una substància exògena, un metabòlit o el producte de la interacció entre l'agent xenobiòtic i una molècula o cèl·lula diana.
- **Biomarcador d'efecte**, que avalua l'alteració bioquímica, fisiològica o de comportament produïda en l'organisme que pot ser associada amb una malaltia.
- **Biomarcadors de susceptibilitat**, és un indicador de la capacitat heretada o adquirida d'un organisme per a respondre a l'exposició a una substància xenobiòtica.

Per poder utilitzar un biomarcador en la pràctica clínica, ha d'haver estat suportat per evidències experimentals. Es requereixen múltiples i llargs estudis amb grans grups d'individus per poder arribar a establir si la presència d'un biomarcador està indicant l'existència d'una malaltia. El procés de selecció i aprovació requereix una acurada consideració de l'especificitat, fiabilitat i sensibilitat d'aquest com a mesura de risc, establint l'exactitud, precisió, a més de la garantia de la qualitat del procediment analític i la interpretació de dades de la medicació, que han de ser comparats amb altres variables.

El biomarcador per tal de ser validat, ha de complir els següents requisits:

- **Identificació del risc**: Confirmar que l'agent és capaç de causar un efecte advers en l'organisme, identificant clarament la condició clínica a avaluar.
- **Avaluació de dosi-resposta**: Establir la relació quantitativa entre dosi i efecte, establint que constitueix un pas relacionat amb la patofisiologia de la malaltia.
- **Avaluació de l'exposició**: Identificar i definir el tipus d'exposició que es produeix, o es preveu que es produeixi.

L'ús de biomarcadors ha d'estar emmarcat en una guia de principis que faciliti el seu ús en la valoració clínica de rutina, la regulació, l'adequada avaluació de la quantificació i validació, tenint en compte les consideracions ètiques necessàries.

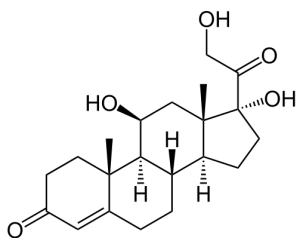
Els factors que requereixen ser considerats en el procés de selecció i validació del biomarcador d'acord amb les directrius de *who-icps 1993* són:

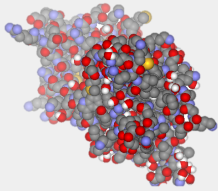
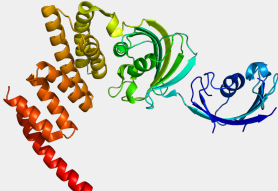
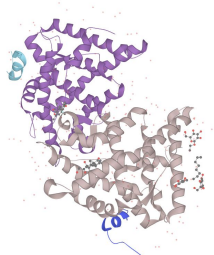
1. Identificació i definició del procés biològic d'interès.
2. Estudis previs de la relació entre l'agent d'exposició, el biomarcador i l'efecte d'avaluar.
3. Identificació de la variable a quantificar, per avaluar la sensibilitat i especificitat del marcador en relació a l'exposició.
4. Selecció d'exàmens disponibles per a l'anàlisi, confiabilitat de la integritat de la mostra entre la recol·lecció i l'anàlisi.
5. Revisió dels procediments analítics disponibles per a la quantificació del marcador i les seves limitacions respecte a la detecció, sensibilitat, precisió i exactitud.
6. Estandarització d'un protocol que garanteix un adequat nivell de qualitat i control.
7. Avaluació de la variació intra i interindividual d'una població no exposada.
8. Anàlisi de dades per establir una relació.

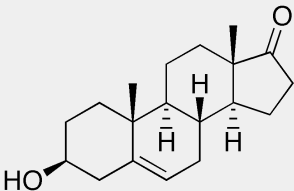
2.2.2. Biomarcadors putatius del PTSD en la literatura

En relació a la literatura del tractament del PTSD, s'examinen biomarcadors fisiològics d'efecte d'entre els quals els més estudiats inclouen mesures de neuroimatge i psicofisiologia, lectures comportamentals i neurocognitives, i analits mesurats en biofluids perifèrics, com la sang i la saliva en el moment inicial o després d'un repte psicològic. Els biomarcadors perifèrics basats en fluids poden incloure indicadors d'inflamació, mediadors de l'eix hipotalàmic, suprarenals, neuroesteroides i neurotransmissors, que tenen rols funcionals tant en el sistema nerviós perifèric com en el central, cosa que possibilita una inferència biològica significativa amb la utilitat clínica. [1]

A continuació es detallen els biomarcadors amb més rellevància en l'estudi del PTSD per la seva freqüència d'aparició en publicacions.

Neurotransmissor	Tipus molecular	Localització	Funcions principals
CORTISOL [9] hidrocortisona 	<i>Hormona glucocorticoide</i>	→ Produït per les glàndules suprarenals.	→ Promou la biosíntesi de glucosa. → Dona resposta a l'estrès. → Combat infeccions. → Regula el nivell de sucre a la sang. → Manté la pressió arterial. → Regula el metabolisme.

BDNF (Brain-derived neurotrophic factor) [10] 	<i>Glucoproteïna</i>	→ Hipocamp → Present en una àmplia gamma de teixits i tipus de cèl.lules, no només al cervell; ○ Retina ○ SNC ○ Neurones motores ○ Ronyons ○ Pròstata.	→ Promou el creixement de connexions neuronals. → Millora el funcionament cerebral, actua com a neurotransmissor i/o ajuda a la comunicació entre cèl.lules diferents. → Estimula la producció de noves cèl.lules cerebrals i enforteix les existents
IL6 [11] 	<i>Glucoproteïna</i>	→ Secretada pels macròfags, cèl.lules T, cèl.lules endotelials i fibroblastos	→ Coordina la resposta del sistema immunològic.
FKBP5 Immunofilina [12] 	<i>Proteïna xaperona</i>	→ Es localitza en múltiples llocs de l'organisme: el cervell, teixit muscular esquelètic i el greix.	→ Regula la sensibilitat del receptor de glucocorticoide (GR) → Es relaciona amb la restricció de l'absorció de glucosa, emmagatzematge de greix i amb l'estrès.
NR3C1 [13] 	<i>Proteïna receptora glucocorticoide</i>	→ Es troba típicament al citoplasma. → Mitjançant la unió a lligand també es transporta al nucli.	→ Codifica el receptor de glucocorticoide → Implicat en respostes inflamatòries, proliferació cel·lular i diferenciació en els teixits diana.

DHEAS (Deshidroepiandrosterona) [14]	<i>Hormona masculina</i>	→ Produïda per la glàndula suprarenal	→ Ajuda a la producció d'altres hormones com la testosterona i l'estrògen.
			

S'han associat les desregulacions neuroendocrines dels nivells de concentració d'aquestes molècules descrites, destacant el BDNF, a un ampli ventall de psicopatologies i/o afectacions neurològiques, com trastorns d'ansietat, depressió, *Alzheimer* o el propi PTSD, tot i que no hi ha consens en si són causa o conseqüència d'aquests trastorns. [2]

2.3. Bases de dades biològiques

Les bases de dades biològiques són sistemes que permeten emmagatzemar, organitzar i recuperar dades biològiques. Aquestes, ofereixen obtenir volums considerables de dades, facilitant el seu intercanvi i la seva recuperació. Tot i que estan restringides a un àmbit d'aplicació molt especialitzat, serveixen com a recursos indispensables als investigadors de diferents disciplines per automatitzar el procés de descobriment del coneixement.

La classificació d'aquestes bases de dades es pot fer de diverses maneres, aquestes inclouen aquelles especialitzades en DNA, RNA, proteïnes, rutes, malalties, ontologies i nomenclatures o categories més àmplies que inclouen bases de dades de seqüència, estructura i rutes. La base de dades desenvolupada en aquest projecte, emmarcant-la en una classificació funcional, es focalitzada en una malaltia, el PTSD, i concretament amb Biomarcadors.

El propòsit de recollir en una mateixa base de dades la informació mencionada sorgeix de la creixent quantitat d'estudis de biomarcadors en totes les àrees de la malaltia, paral·lel al creixent nombre de metanàlisis que combinen dades de múltiples estudis per obtenir conclusions comunes sistemàticament. D'aquesta manera, es fa patent la manca d'harmonització dels resultats d'aquests estudis, que sovint romanen en forma de text no estructurat, xifres, taules o fitxers complementaris. Organitzar i emmagatzemar aquest coneixement és essencial per fomentar el descobriment i el desenvolupament de diagnòstics i tractaments.

Existeixen bases de dades, complementades amb una plataforma web, amb un enfoc similar al que s'ofereix en aquest projecte. És el cas de la data base *PTSDpubs* [15], desenvolupada i

mantinguda pel Centre Nacional Americà pel PTSD. Aquesta proporciona accés a la literatura del PTSD i altres conseqüències sobre la salut mental de l'exposició a esdeveniments traumàtics. No obstant, no rep l'enfoc de centrar-se amb els biomarcadors, sinó simplement compilar totes les publicacions referents al PTSD sense establir una estructuració del seu contingut, tret diferencial de PTSDDDB i on s'explota la necessitat.

3. Objectius

Aquesta memòria, en base a la formulació del projecte PTSDDDB donada la necessitat de recopilar i consensuar en una mateixa base de coneixement les consideracions i dades recollides la literatura sobre els biomarcadors del PTSD, estableix els següents objectius:

- Revisar la tasca de curació manual realitzada per popular la base de dades.
- Detallar els procediments de modelatge de la base de dades i de desenvolupament de la plataforma web.
- Exposar de forma descriptiva els components de la plataforma PTSDDDB.
- Utilitzar les eines que ofereix l'aplicatiu web per accedir a les dades públiques, validant la seva usabilitat.
- Realitzar un anàlisi de dades per determinar el consens en la literatura dels biomarcadors putatius.

4. Materials i Metodologia

En aquest apartat es revisen les metodologies i eines emprades en les diferents fases de desenvolupament del projecte, examinant primer la confecció de la base de dades de PTSD i a continuació els mètodes d'anàlisi de dades aplicats, i que s'esquematitza en la figura següent:



Figura 1. *Esquema de les fases de desenvolupament de PTSDDDB.*

4.1. Procediment de curació

Com a punt d'arrencada del projecte, es va determinar i reunir la informació a extreure de la literatura, tot integrant-la en un format portable i estructurat amb la finalitat de popular la base de dades. Per dur-ho a terme, s'ha realitzat una extensa tasca de curació manual en la qual s'han recol·lectat dades i/o anotacions a partir de la lectura i extracció acurada del contingut exposat en literatura. A continuació es detallen les metodologies i procediments seguits en aquest procés.

Selecció de corpus

Per seleccionar les publicacions, s'han seguit dos criteris d'inclusió:

- Recomanacions i referències d'experts en el camp.
- Mesures de mineria de dades, tenint en compte el numero referències citades en publicacions de revisió.

Les publicacions incloses, consten d'articles originals publicats entre 1997 i 2018, garantint l'actualitat de la recerca, els quals han examinat biomarcadors basats en fluids en humans centrant-se en pacients amb PTSD, tot comparant-los amb poblacions de control (controls sans, controls exposats a traumes i/o pacients amb trastorns psiquiàtrics, entre altres comorbiditats).

El criteri d'exclusió principal ha demarcat publicacions que no integren poblacions amb PTSD, d'aquelles que comptaven amb pacients amb PTSD, però en absència de biomarcadors basats en fluids, o que eren estudis preclínic.

Finalment, derivat del procés de selecció, s'han inclòs 109 articles.

Extracció de dades

La informació continguda en els 109 articles seleccionats, va ser extreta de forma independent per cinc curadors formats i afegida a una plantilla de model de dades. El treball manual d'aquests ha consistit en examinar amb detall les publicacions i introduir la informació en diferents entrades de manera paral·lela, per després arribar a un consens de les diferents extraccions.

L'esquema de metadades s'ha anat construint de forma incremental, partint d'un model de deu articles i afegint nous camps de metadades en funció de les consideracions realitzades en l'anàlisi de manuscrits addicionals. L'espectre de metadades cobreix molts camps, incloent el disseny de l'estudi, la demografia, els resultats de l'estudi, la informació de l'assaig i els mètodes estadístics, com es llista amb més detall en el model conceptual final de la base de dades. Cal mencionar que degut a l'heterogeneïtat d'informació, els camps de la plantilla s'interrelacionen de manera complexa i les entrades poden contenir diversos camps nuls, ja que els diferents estudis no cobreixen la mateixa informació.

Avaluació de la qualitat

S'han realitzat tres rondes de control de qualitat per garantir la validesa de les metadades. A cada ronda de control de qualitat, les metadades van ser revisades per un curador diferent; si s'identificaven incongruències, els curadors arribaven a un acord. A més, el procediment de control de qualitat generava l'increment de més metadades, alhora que incloïa més publicacions.

Per altra banda, es va implementar un analitzador de la plantilla de curació en base a la població del model la base de dades, que controlava de forma automatitzada la qualitat del full de treball mitjançant la identificació de duplicats, la comprovació de la sintaxi i la normalització dels termes.

4.2. Model de la base de dades

Un dels desafiaments en la integració de dades a l'hora de gestionar fonts d'informació dispars, és organitzar el coneixement en un esquema comú. La determinació d'aquest patró d'integració, influeix considerablement tant en la recollida de dades, en els passos posteriors de gestió i població de la base de dades, com en l'eficàcia de la mostra final d'informació i l'anàlisi de dades. Atès que es van acabar incloent un gran nombre de variables i metadades, en un ordre d'interrelació complex, i aquestes dades són sovint amb un format no estructurat, va suposar una tasca àrdua establir un modelatge integratiu i que habilités la presència de camps nuls/buits.

Per estructurar la informació que es recull a les 193 columnes/camps de la plantilla de curació, s'ha dissenyat un esquema de dades relacional, el qual emmagatzema aquesta informació en 17 models diferents en base als diferents atributs integrats (per exemple, publicació, biomarcador, estudi clínic, metadades de cohorts, etc.), representats com a taules a la base de dades. A continuació es descriu cada mòdul/taula del model relacional, llistant els seus camps rellevants. (*Consultar a annex 3 l'esquema relacional de la base de dades*).

Donat que gran part de la metadata és de caire mèdic, s'empra nomenclatura corresponent a estudis clínics, de la qual en la següent taula es fa una breu descripció per tal de contextualitzar-la, tot relacionant-la amb continguts mencionats en la descripció de l'enfermetat (*Apartat 2.1*). [16]

Model/Taula	Camps	Descripció
Biomarcador	- Nom de la publicació - Nom estandaritzat - Indicació primària	El nom estandaritzat té la finalitat d'harmonitzar les entrades, si per un biomarcador en concret es nomenen de manera diferent a l'estudi, s'han de guardar com una mateixa entrada. Les nomenclatures seguides han estat: - HUGO ID o qualsevol altre acrònim o símbol del gen. - HGNC en els biomarcadors de proteïnes, gens o miRNA que es codifiquen utilitzant la nomenclatura si és possible. - ChEBI, PubChem i InChIKeys. Per molècules petites, prioritant-se per l'ordre establert.
Indicació	- Nom - Detalls malaltia	En medicina, una indicació es coneix com a signe, símptoma o afecció mèdica que

		conduïx a la recomanació d'un tractament, prova o procediment. En PTSDDDB es mostren indicacions com: trastorn per estrès posttraumàtic, trauma infantil, PTSD matern.
Estudi	- Identificador <i>PubMed</i> - Informació d'autoria: - Noms - Any de publicació - Informació geogràfica	Informació descriptiva de la literatura.
Criteris d'inclusió/exclusió	- Controls actius - Criteri d'exclusió - Criteri d'inclusió - Malaltia mental terminal - Medicació	Els criteris d'inclusió/exclusió delimiten la selecció dels pacients.
Punt de temps de l'estudi (<i>time-point</i>)	- Time-point	Temps relatiu en que s'ha realitzat l'estudi/mesura, pot ser dos setmanes després del trauma, dos anys després del trauma, etc.
Resum numèric	Mesures: - Mitjana - Desviació estàndard. Per a cada: - Indicació primària - Controls exposats a trauma - Controls saludables.	Estadístiques de les concentracions per als diferents grups de l'assaig d'un biomarcador en concret.
Instrumentació clínica	Instrumentació clínica per: - Indicació primària - Trauma en adults - Infància - Vida útil	Informació descriptiva de la instrumentació clínica en cada grup del cohort.
Estudi clínic	- Tipus d'estudi - Estudi de resposta al tractament - Nombre d'individus per indicació	Especifica els detalls de l'estudi, podent ser longitudinal, transactional.
Comorbilitat	- Nom - Mesura de comorbilitat - Nombre Mesures: - Mitjana	Mesures de comorbilitat, dotant amb informació numèrica el grau d'interrelacions entre patologies.

	- Desviació estàndard. Per a cada: - Indicació primària - Controls exposats a trauma - Controls saludables - Altres psicopatologies	
Nom de cohort i detalls demogràfics	Mesures: - Mitjana - Desviació estàndard - Percentatge de dones - Ascendència(caucàsica, afroamericana, etc.) Per a cada: - Indicació primària - Controls exposats a trauma - Controls saludables - Altres psicopatologies	Dades i detalls descriptius del cohort, és a dir del grup d'individus en el que s'ha aplicat l'estudi.
Resultats Direcció del canvi	Direcció del canvi: - Incrementat - Decrementat - No canvi - Canvis de circuits específics. - Notes - Descripcions	La direccionalitat del canvi del biomarcador en una evidència en particular sota l'efecte d'una indicació clínica.
Assaig	- Marca - Sonda - Fluid - Substrat - Detecció de límits d'assaig - Unitats de mesura	Detalls instrumentals de l'assaig.
Càlculs d'assaig	- Mitjana - Desviació estàndard. - Mida de l'efecte - Combinació Per a cada: - Indicació primària - Controls exposats a trauma - Controls saludables - Altres psicopatologies	Mesures associades a un assaig

Informació estadística	- Mitjana - SD - Expressió fold change - Variància - Canvi de metilació - t valor - p valor - q valor	Aquesta informació es complementa incrementalment amb l'adició de dades suplementàries als estudis, més allà del protocol de curació, aquestes s'han importat programàticament. Precisament aquests valors permeten dotar d'informació quantitativa suficient a les entitats.
-------------------------------	---	---

4.3. Desenvolupament de l'aplicatiu web

Per tal d'implementar el model relacional descrit de la base de dades, popular les dades a emmagatzemar i visualitzar-ho en format web, s'han emprat procediments i eines de desenvolupament de programari, seguint una arquitectura de programari Model Vista-Controlador (MVC). Aquest model divideix l'aplicació en tres parts interconnectades: el model/base de dades, la interfície d'usuari i la lògica de control.

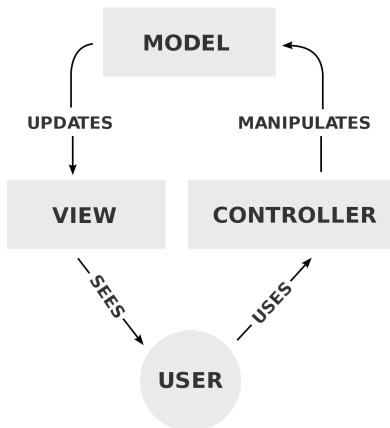


Figura 2. Esquema del Model Vista-Controlador (MVC).

4.3.1. Back-end, modelatge i gestió de la BD: Django i Python

El codi de back-end s'escriu en *Python*, tot suportant-se sobre la infraestructura *Django*. [17] *Django* és un *framework* web Python d'alt nivell que proporciona un desenvolupament ràpid amb un disseny pràctic abraçant-se al paradigma MVC. Això és possible gràcies a que grans seccions de codi que implementen el MVC, són proporcionades pel *framework*, i el desenvolupador es centra en 'emplenar' les parts que generen la seva aplicació en particular. És lliure i de codi obert, amb una comunitat pròspera i proactiva, una documentació acurada i moltes opcions de suport gratuït. Les dades s'emmagatzemen en una base de dades relacional MySQL controlada per vistes, permetent fer consultes dirigides, oferint els continguts accessibles programàticament als desenvolupadors per a la seva manipulació.

La població del model descrit es fa mitjançant la càrrega de la plantilla de curació, executant un analitzador/parser que omple els models *Django*, emmagatzemant la informació carregada a la base de dades MySQL. Aquest procés no només permet plenar la base de dades per realitzar consultes posterior, sinó també comprovar la qualitat del full de treball i realitzar la normalització de les dades.

4.3.2. Front-end, interfícies: Javascript i OpenAPI

Com que l'objectiu principal de l'aplicació web és l'exploració i la visualització de dades, en el front-end s'han implementat diferents visualitzacions interactives mitjançant biblioteques de *Javascript*: *D3.js*, *C3.js*, *DataTables* i *DataMaps*. A més es complementa per *Bootstrap*, garantint un disseny gràfic harmonitzat i conservant total compatibilitat amb diferents dispositius (*Smartphones*, tablets, ordinadors portàtils, etc.).

Finalment, PTSDDDB es complementa amb una API *REST* [18] documentada amb especificació *OpenAPI*. Una API *RESTful* es tracta d'una interfície de programa d'aplicació (API), aprofitant les metodologies HTTP. En el cas de l'API PTSDDDB solament és proporcionen *endpoints* amb sol·licituds GET, ja que no es desitja la modificació de la base de dades sinó solament proporcionar les dades en un format portable programaticament.

4.4. Mesura del consens: Entropia de *Shannon*

Per determinar el nivell de consens entre els estudis d'un biomarker determinat s'utilitza l'entropia de *Shannon*. Aquesta mètrica mesura la taxa mitjana amb la qual la informació és produïda per una font de dades estocàstica, és a dir, que sigui fruit de l'aleatorietat. Aquesta es calcula amb la següent funció de probabilitat: [19]

$$S = - \sum_i P_i \log P_i.$$

D'aquesta manera, l'entropia de *Shannon* pot ser utilitzada com una mètrica del nivell de concordança entre un conjunt d'informacions, pel qual s'utilitza com a indicador del consens entre estudis. Dit d'un altra forma, donades evidències en diferents publicacions dels efectes, l'entropia mesura el 'desordre' o 'desconsens' en la literatura, o com es presenten efectes contradictoris. Així, a nivells baixos d'entropia, el consens entre els estudis s'identifica major.

5. Resultats

En el següent apartat s'exposen els resultats del projecte, primer detallant el contingut final de la plataforma web i a continuació discutint la primera aproximació d'anàlisi sobre les dades obtingudes mitjançant consultes dirigides al contingut de PTSDDDB.

5.1. Aplicatiu web PTSDDDB i descriptiva de les dades

Com ja s'ha mencionat, les dades públiques del projecte, es distribueixen a través de la plataforma web PTSDDDB (<https://ptsd.scai.fraunhofer.de/>), la qual proporciona diferents interfícies interactives que faciliten l'exploració i l'intercanvi d'informació dels biomarcadors catalogats. Cal destacar que la plataforma pública no ofereix totes les dades que la base de dades inclou, solament les que s'han considerat de domini públic; les que s'han considerat reservar ha estat tant per motius de valor de les pròpies, com per estar pendent de control de qualitat. En el següent apartat es descriuen les diferents planes de la plataforma web amb els components que mostren/contenen.

Plana principal i d'exploració

S'exposa de forma visual i accessible una breu descripció i l'enllaç a les planes públiques de la plataforma web a través d'un *Carousel*. A més, la barra d'exploració superior, ordena i proporciona un accés ràpid a les planes web des de qualsevol punt de l'aplicatiu.

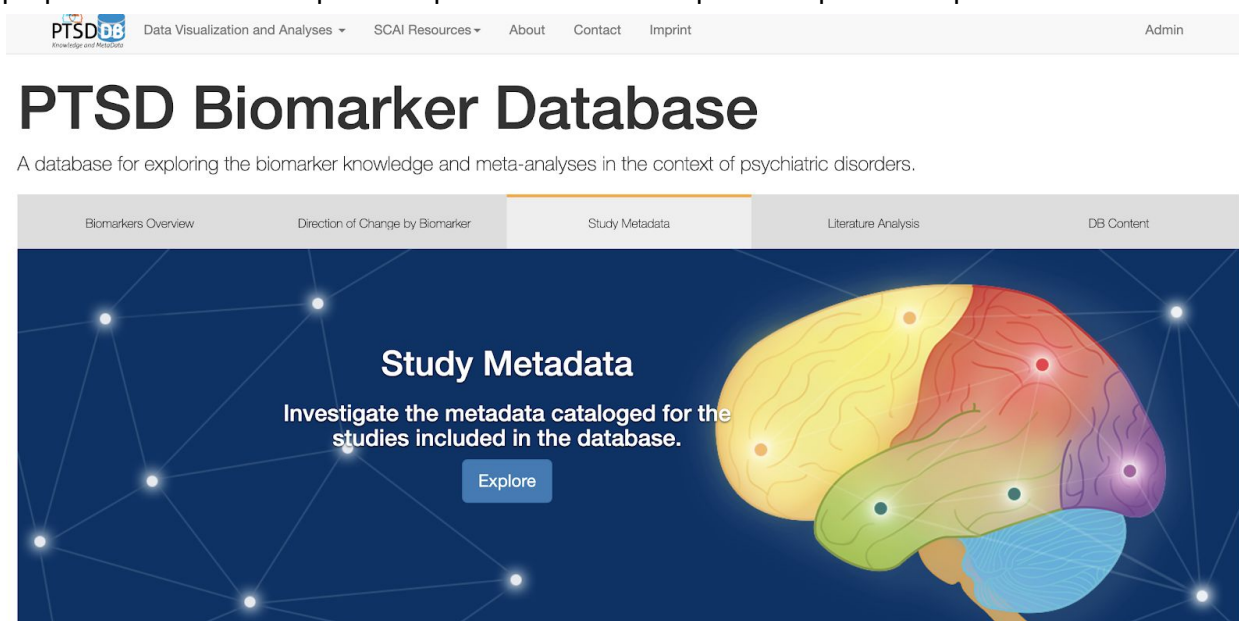


Figura 3. Mostra del *Carousel* a la pàgina principal.

5.1.1. Resum de freqüències de biomarcadors

<https://ptsd.scai.fraunhofer.de/frequencies>

La plana web recull estadístiques descriptives del còmput de publicacions en que han estat estudiats els biomarcadors catalogats. Es mostren en gràfics de barres, en primer lloc les **freqüències globals**, i a continuació **classificats pels biofluids** en els que se'ls ha analitzat i pel tipus de substrat. Aquests proporcionen una visió de la dimensió de la base de dades, així com una idea global de la rellevància dels biomarcadors en funció a la freqüència de com d'estudiats/citats han estat. Si es passa per sobre la barra s'ofereix més informació del biomarcador i si es clica es mostren els links a les entrades dels articles a *Pubmed*.

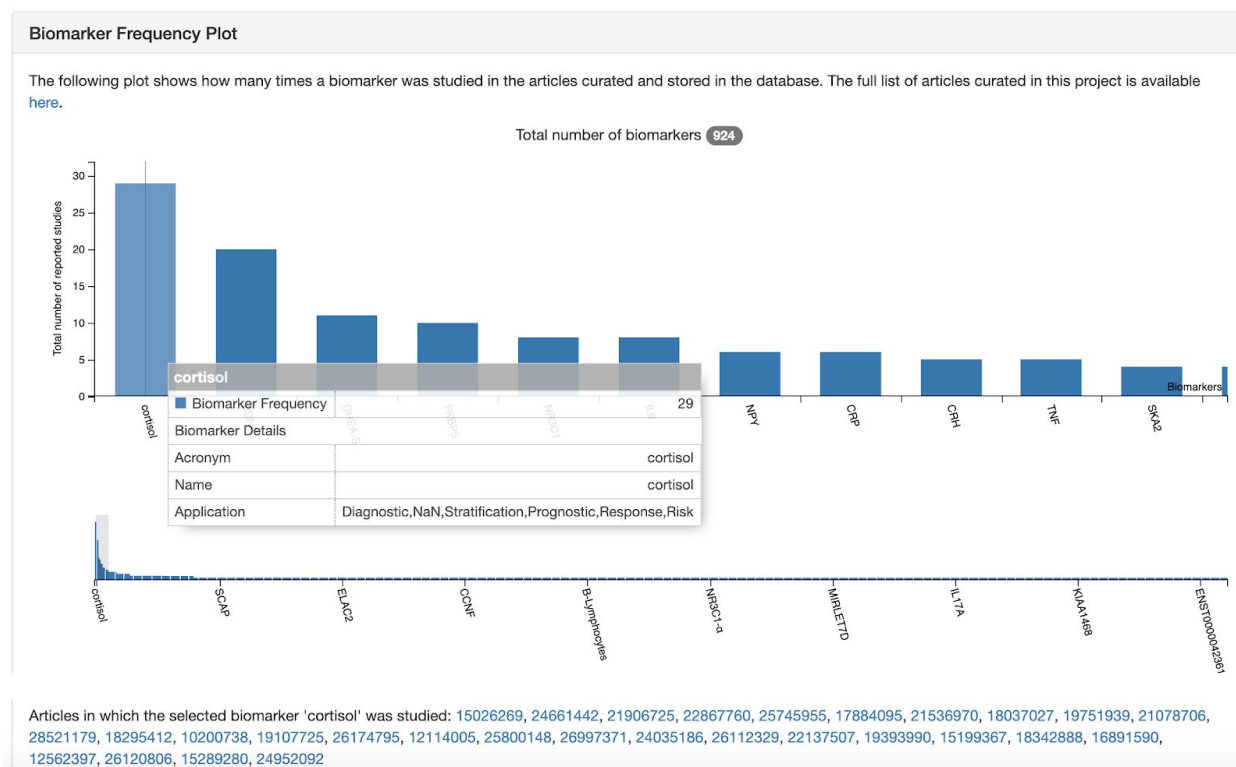


Figura 4. *Freqüències globals dels biomarcadors a la literatura examinada.*

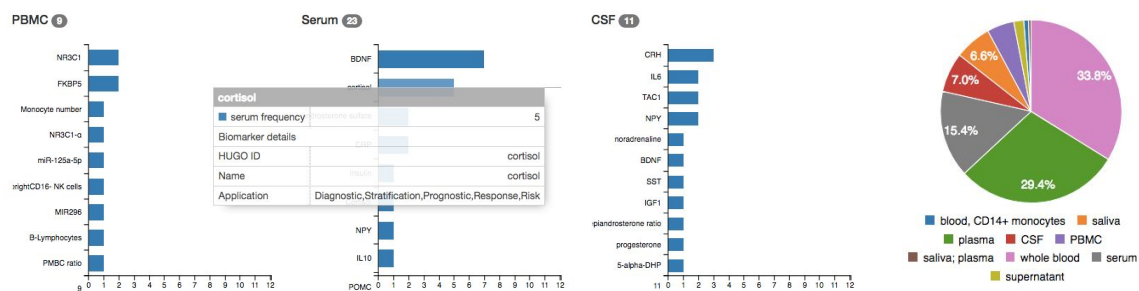


Figura 5. *Freqüències de publicacions dels biomarcadors per biofluids en què es van mesurar i una representació de la proporció de mesures per a cada biofluids.*

5.1.2. Direcció del canvi relatiu per biomarcador

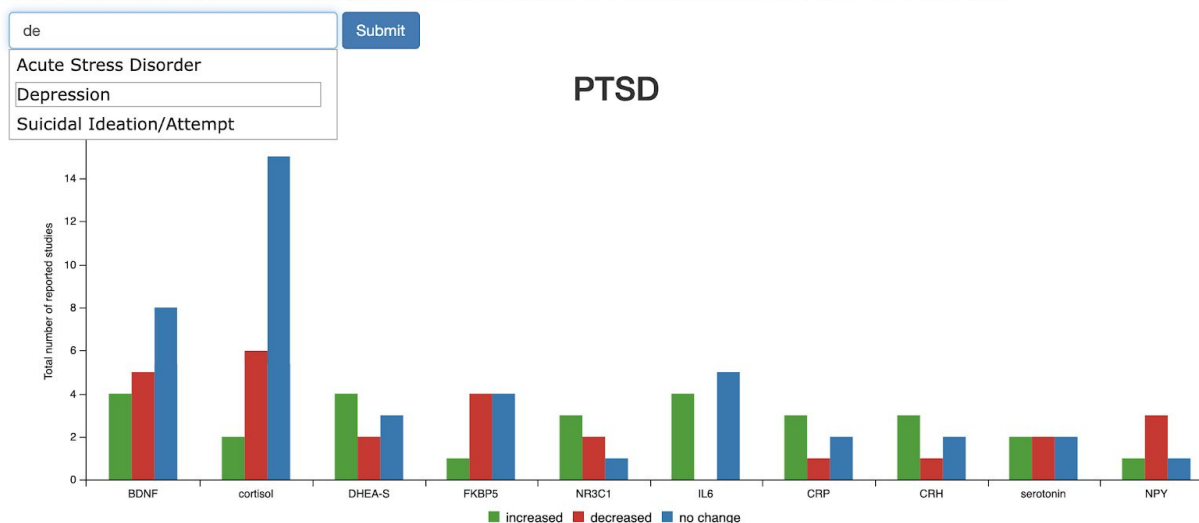
https://ptsd.scai.fraunhofer.de/relative_changes

En aquesta secció de l'aplicatiu web, es visualitzen les diferents **evidències del canvi relatiu** d'un biomarcador **sota l'efecte d'una indicació** (patologia i/o símptoma) determinada. És a dir, es mostren les diferents freqüències en les que s'ha demostrat en la literatura la direcció del canvi relatiu en la concentració d'un biomarcador sota una indicació clínica respecte els controls; si en presència del símptomes, l'expressió del biomarcador ha incrementat, ha decremuntat o no ha produït cap canvi (efecte nul). Això permet examinar els nivells de discrepància dels resultats d'estudis. Es presenten dues visualitzacions: un gràfic de barres de tres dimensions i una taula descriptiva classificada pel tipus de substrat.

Histograma de la direcció del canvi

Es mostren, donada una indicació clínica (per defecte es mostra PTSD), com a màxim els deu primers biomarcadors per la freqüència d'articles en els que es reporta que en presència de la patologia s'ha evidenciat un canvi relatiu (increased, decreased i/o no change) representant les evidències contradictòries que es troben a la literatura. Cada una de les freqüències dels canvis relatius es mostra en base a diferents colors de barres. Els usuaris poden seleccionar altres indicacions (com depressió, estrès, intent o ideació de suïcidi, etc.) a partir d'un formulari de cerca, i clicant sobre una barra en concret, obtenir el llistat d'enllaços als articles identificadors de *PubMed* que suporten l'evidència en qüestió.

The chart below contains the relative changes for the most common biomarkers collected in PTSD. You can also explore the relative changes for the most common biomarkers in other conditions by using the form below. Finally, at the bottom of the page, you can find the relative changes observed in a specific biomarker.



Articles in which 'BDNF was found INCREASED: 20097247, 25706090, 23774996, 23319005

List of biomarkers available in this indication (they can be queried in the form at the bottom of the page):

BDNF, cortisol, DHEA-S, NR3C1, FKBP5, NPY, CRP, CRH, serotonin, IL6, TNF, Diastolic blood pressure, IFNG, CCL5, DDX17, TMEM167A, 2-arachidonoylglycerol, progesterone,

Figure 6. Mostra de la visualització per a la indicació de PTSD, els diferents nivells de còmput d'articles de com a màxim deu primers biomarcadors per direcció de l'efecte.

Taula per substrat biològic

Per contextualitzar millor la direcció del canvi, un altra secció de la plana (accessible per una barra de pestanyes) proporciona als usuaris la funcionalitat d'avaluar els canvis basats en el substrat biològic d'on va ser mesurat el biomarcador. Un quadre interactiu presenta els resultats numèrics del biomarcador per a diferents substrats biològics (RNA, metabolit, proteïnes, etc.). En aquest cas es navega a la barra de cerca per biomarcador, mostrant els valors totals per a totes les indicacions sense seleccionar-los per aquestes.

Relative Change Found in Each Biomarker by Indication

Direction of Change by Biological Substrate

Direction of Change by Biological Substrate

Below, relative changes for a given biomarker in different biological substrates can be queried. The example shows the observed relative changes for BDNF.

BDNF

Substrate	Increased	Decreased	No change	Valence
DNA methylation	↑ 2	↓ 1		+ 1
Protein	↑ 5	↓ 4	= 8	+ 1
DNA			= 1	
RNA		↓ 1		- 1

Figura 7. Direcció del canvi per substrat biològic per a un determinat biomarcador.

5.1.3. Anàlisi de la literatura

<https://ptsd.scai.fraunhofer.de/literature>

Aquesta plana de l'aplicatiu web es centra en presentar dades descriptives de la biografia inclosa a la base de dades, mitjançant visualitzacions de dades que assenyalen on i quan s'han publicat els articles i exposen els punts més o menys investigats del paisatge clínic. Això proporciona a l'investigador la possibilitat de proposar enfocaments potencials d'investigació, tenint en compte quins biomarcadors ja s'han estudiat i en quin enfocament, i també pot fomentar la col·laboració entre investigadors que treballen en hipòtesis similars.

Taula descriptiva

En primer lloc, una taula interactiva llista la informació identificativa dels articles: ID de PubMed (amb enllaç), títol, diari, autors i any de publicació. Es possibilita ordenar els articles pels diferents camps i realitzar cerques dirigides a través d'un formulari.

Articles included in PTSDDB					
Details about the articles included in PTSDDB.					
Show	10	entries	Search:		
PubMed ID	Title	Journal	Publication Year	Authors	
9137116	Elevated CSF corticotropin-releasing factor concentrations in posttraumatic stress disorder.	Am J Psychiatry	1997	Bremner JD,Licinio J,Darnell A,Krystal JH,Owens MJ,Southwick SM,Nemeroff CB,Charney DS	
10200738	Serial CSF corticotropin-releasing hormone levels and adrenocortical activity in combat veterans with posttraumatic stress disorder.	Am J Psychiatry	1999	Baker DG,West SA,Nicholson WE,Ekhator NN,Kasckow JW,Hill KK,Bruce AB,Orth DN,Geraciotti TD Jr	
10202571	Low platelet-poor plasma concentrations of serotonin in patients with combat-related posttraumatic stress disorder.	Biol Psychiatry	1999	Spivak B,Vered Y,Graff E,Blum I,Mester R,Weizman A	
10715359	Low baseline and yohimbine-stimulated plasma neuropeptide Y (NPY) levels in combat-related PTSD.	Biol Psychiatry	2000	Rasmusson AM,Hauger RL,Morgan CA,Bremner JD,Charney DS,Southwick SM	

Figura 8. Taula d'entrades descriptives de la literatura continguda a PTSDDB.

Distribució geogràfica

Una visualització de mapes interactius visualitzen la distribució geogràfica dels articles curats permetent identificar centres de concentració d'investigació de marcadors de PTSD als Estats Units i a tot el món, la qual cosa pot ajudar a identificar oportunitats de col·laboració per a la replicació i validació de biomarcadors. Es pot observar la concentració de publicacions als Estats Units, amb menor mesura a la Xina, i a Europa destaquen Alemanya i Suècia.

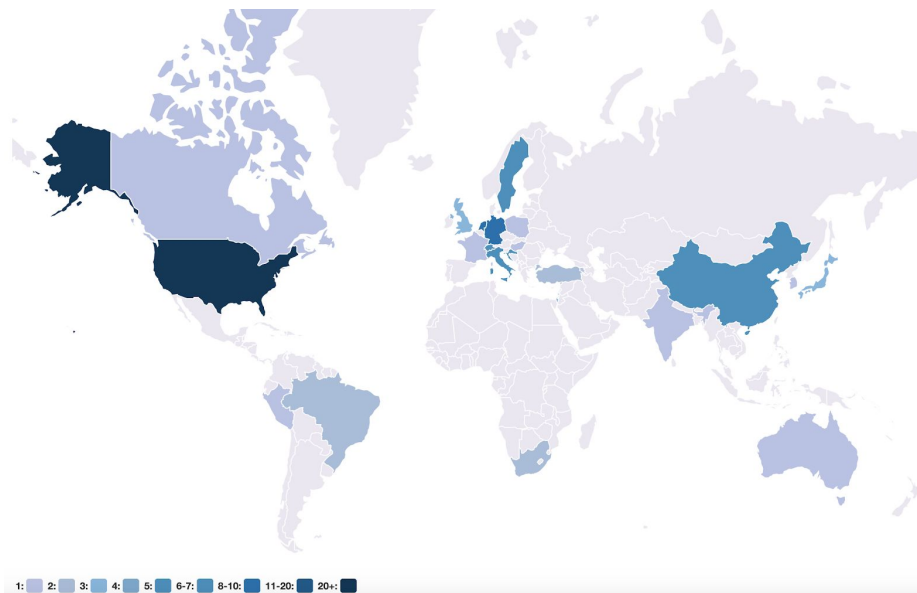


Figura 9. Visualització interactiva de la distribució geogràfica mundial per països de les publicacions. Si es clica sobre un país s'obté la llista d'identificadors amb l'enllaç a Pubmed.

Cronologia de publicacions

Histograma que il·lustra els anys de publicació dels articles inclosos. Aquest anàlisi és de rellevància ja que la investigació de biomarcadors sovint es basa en les tendències, les tecnologies i les hipòtesis específiques del domini temporal de la recerca. Cal per tant prioritzar els articles actuals, dels quals és dels que més freqüència de literatura se'n disposa (dels últims cinc anys).

Analysis of the years of publication of the studies included in PTSDDDB.

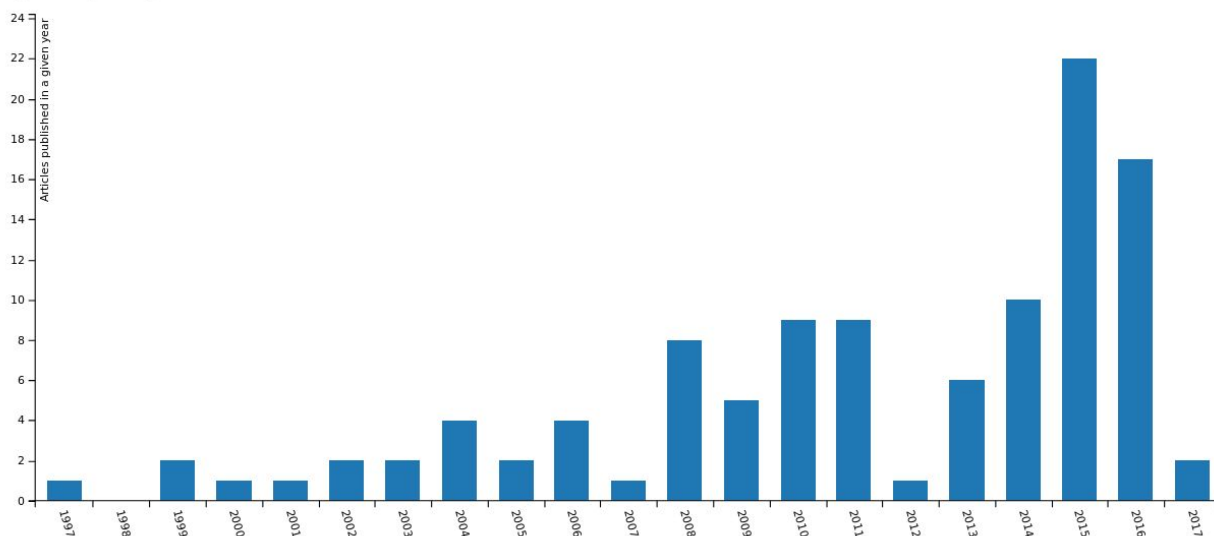


Figura 10. Anys de publicació dels articles inclosos a PTSDDDB.

Co-ocurrència

En la literatura de biomarcadors, normalment s'estudien de manera conjunta parelles d'aquests degut a que estan implicats en processos regulatoris vinculats. L'última visualització de la plana de literatura, il·lustra la freqüència en què combinacions de biomarcadors s'estudien junts en la mateixa publicació mitjançant mapes de calor; donant una idea de la co-ocurrència i localitzant buits potencials per examinar en investigacions posteriors.

A més, s'han implementat dos mapes de calor complementaris per explorar la co-ocurrència en en la combinació de substrats biològics utilitzats en els estudis extrets i en les combinacions de substrats biològics i biofluids utilitzats.

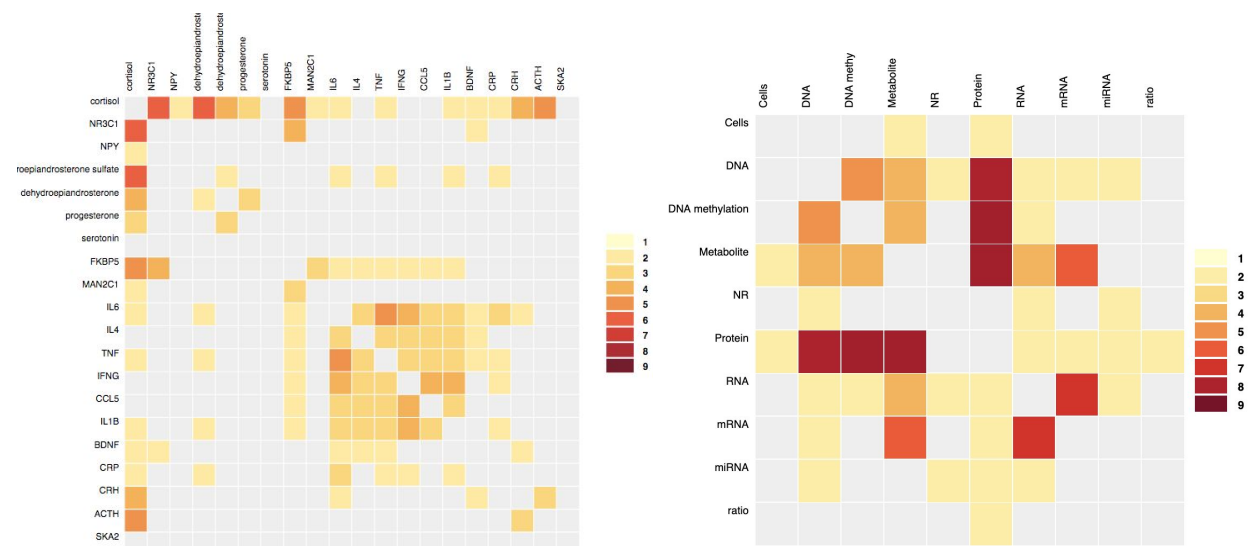


Figura 11.. Es pot interactuar amb el mapa de calor permetent agrupar i ordenar biomarcadors; en l'exemple, s'ordena pel nombre total de publicacions, llistant primer el cortisol que s'estableix com el biomarcador més estudiat.

5.1.4. Metadades dels estudis

<https://ptsd.scai.fraunhofer.de/study>

Per explorar les metadades clíniques recopilades dels estudis inclosos al PTSDDB, s'ofereix una plana amb taules interactives. Aquestes metadades consten d'anotacions i dades dels assajos clínics que han avaluat el biomarcador consultat, les quals més que informació d'evidències, proporcionen informació procedimental i instrumentals (mida de la mostra, tipus d'estudi, durada, diagnòstic, mesures estadístiques: mitjana, desviació estàndard, etc.), tot ordenant la informació sota un esquema comú i oferint-la als investigadors per a la seva examinació.

La disponibilitat i integració de metadata clínica és crucial per comparar diversos dissenys d'estudi que són sovint de naturalesa complexa i heterogènia, oferint la possibilitat de desenvolupar estudis com models predictius o, anàlisis longitudinals en base a les seves anotacions.

S'han seleccionat un total de 18 camps de metadades a fer públiques a través d'una taula interactiva. Mitjançant el formulari es permet navegar per la informació, filtrant la consulta en base a la selecció d'un biomarcador i una certa aplicació clínica (diagnòstic, pronòstic, risc, etc.). Per últim, un quadre de cerca ràpida llista els biomarcadors analitzats en un estudi concret per tal d'enllaçar la metadata amb la resta de planes de PTSDDB.

Study Metadata

This page allows users to explore the collected metadata in the studies included in PTSDDDB. The form below, presents information about the studies containing the queried biomarkers. Since the table that contains the information has 14 columns and cannot be visualized completely, users can scroll over the table.

BDNF

Diagnostic

Submit

Clinical studies metadata containing 'BDNF' as a biomarker

Show

10

entries

Search:

PubMed ID	Type	Timeline	Challenge type	Treatment response	Primary Indication	Trauma exposed controls	Other psych subj tested	Other psych subj	Clean controls	Other psych patients indication	Extra notes primary indication
19409951	Cross-sectional	NaN	NaN	False	18	NaN	NaN	NaN	18	NaN	NaN
20097247	Cross-sectional	NaN	NaN	NaN	34	NaN	NaN	NaN	34	NaN	Trauma-exposed patients include acute and post traumatic stress w/ recent (<1 year) or remote (>4 years) trauma exposure
20097247	Cross-sectional	NaN	NaN	NaN	13	NaN	NaN	NaN	34	NaN	Trauma-exposed patients include acute and post traumatic stress w/ recent (<1 year) or remote (>4 years) trauma exposure

Figura 12. Taula d'exploració de les metadades amb el formulari de filtratge. Les 18 columnes no es poden visualitzar completament, pel qual els usuaris poden desplaçar-se sobre aquesta.

5.1.5. Accés al contingut de la base de dades: API RestFUL

<https://ptsd.scai.fraunhofer.de/swagger-ui>

L'API restufull implementada com a part de l'aplicatiu web de PTSDDDB ofereix la possibilitat de consultar i obtenir fàcilment dades contingudes en l'aplicatiu en un format portable per a la seva manipulació. Aquesta permet a investigadors interessats en models de dades recuperar dades a les quals no es pot accedir mitjançant les visualitzacions interactives, ja que tot i presentar-les en un format visual i interactiu no es poden 'importar' en un format apte per a la seva manipulació, a part que no tota la informació continguda a PTSDDDB es presentada en les visualitzacions presentades anteriorment.

La interfície d'usuari de PTSDDDB Swagger ofereix una visualització interactiva dels *endpoints* disponibles en l'API de PTSD, permetent als usuaris desenvolupadors i consumidors interactuar amb aquests i construir consultes dirigides, retornant les dades amb una estructura a ser manipulada programàticament.

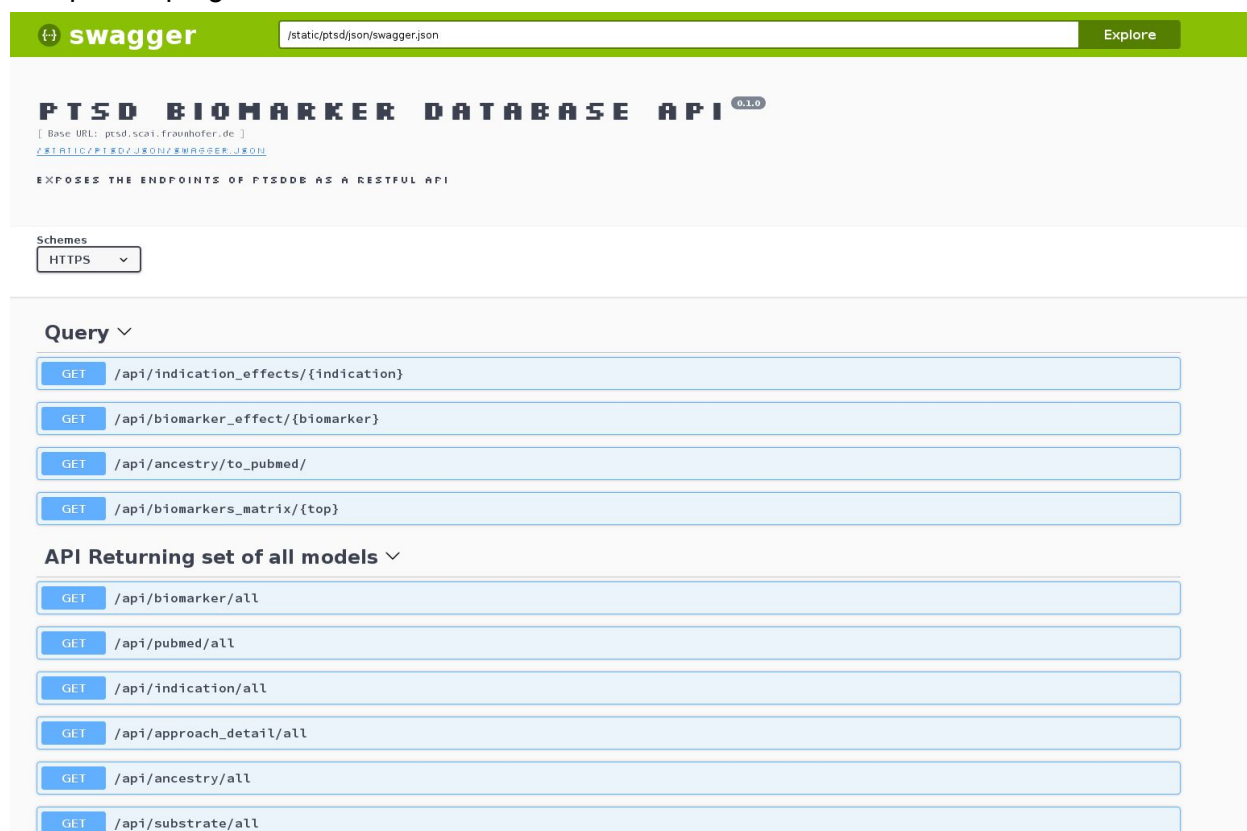


Figura 13. Interfície d'usuari de PTSDDDB Swagger la qual ofereix una visualització interactiva del llistat d'endpoints de l'API de PTSD.

5.2. Anàlisis de dades de PTSDDDB

En aquest apartat final de resultats, es mostra la primera proposta d'anàlisis sobre el contingut de PTSDDDB, més enllà de la informació descriptiva que contenen les planes públiques de la interfície web. Es pot consultar el codi implementat en un *Jupyter Notebook*, per dur a terme aquests anàlisis a l'annex 4.

5.2.1. Consultes dirigides mitjançant l'API de PTSDDDB

Com s'ha exposat en l'*apartat 5.1.4*, l'API implementada com a part de l'aplicatiu web de PTSDDDB proporciona la possibilitat de consultar i retornar fàcilment dades contingudes en base de dades en un format portable per a la seva manipulació.

D'aquesta manera, complint l'objectiu de validar la usabilitat de les eines de l'aplicatiu web desenvolupat, es fa ús d'aquesta, construint consultes dirigides mitjançant els *endpoints* de l'API, per obtenir el conjunt de dades necessàries a aplicar en els anàlisis proposats en aquesta tesis.

5.2.2. Substrats dels biomarcadors putatius més freqüents

Com ja s'ha mencionat, un criteri principal per correlacionar un biomarcador amb un efecte/causa és considerar la freqüència d'estudis en què s'ha estudiat la relació del biomarcador a la patologia. En aquest cas, es presenten els cinc biomarcadors més freqüentment associats al PTSDDDB, la caracterització bioquímica de la qual s'ha descrit en l'estat de l'art, classificant el substrat en el que se'ls ha estudiat (metabolit, proteïna o RNA).

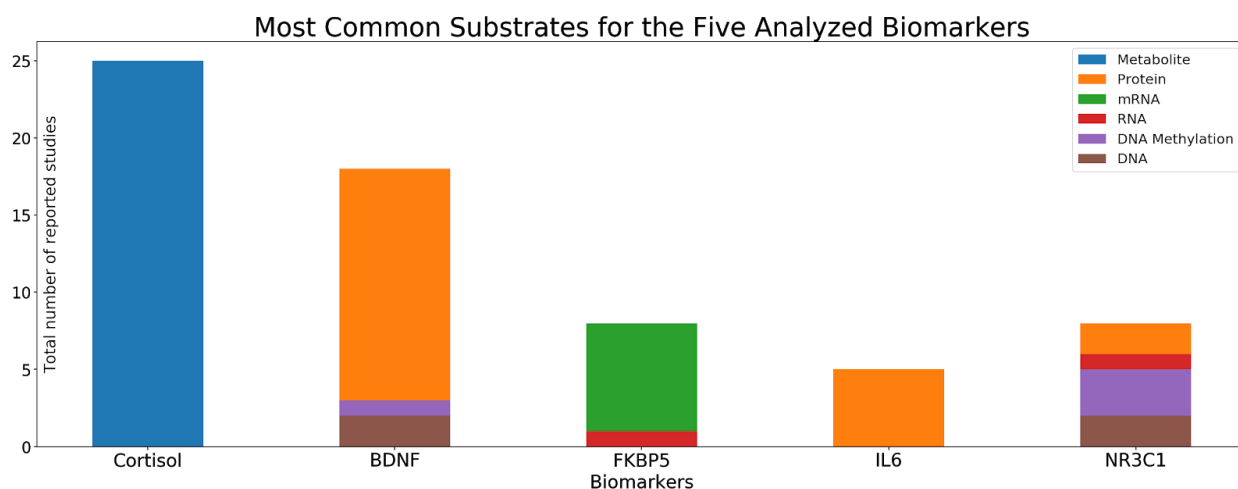


Figura 14. Freqüències dels biomarcadors estratificats per substrats.

5.2.3. Anàlisi del consens en el canvi relatiu dels biomarcadors

Amb la finalitat de determinar i seleccionar els biomarcadors que mostren més nivell de consens en la literatura, tenint en compte les evidències de canvi relatiu dels biomarcadors individuals sota l'efecte del PTSD (*apartat 5.1.2*), es realitza un anàlisi en base les freqüències de les evidències per a cada biomarcador.

5.2.3.1. Càlcul de l'entropia de *Shannon*

S'aplica el càlcul de l'entropia prèviament descrit amb les freqüències de la direcció de canvi relatiu en l'efecte del PTSD (*increased, decreased, no-change*) per cada biomarcador.

A continuació s'exemplifica el càlcul per a BDNF, donat el resultat de la consulta obtingut amb l'URL https://ptsd.scai.fraunhofer.de/api/biomarker_effect/BDNF de l'API:

```
{
  "Increased": "8",
  "Decreased": "6",
  "No change": "8"
}
```

S'obté el còmput total d'evidències:

$$total\ evidències = \Sigma\ evidències = increased + decreased + no\ change = 8 + 6 + 8 = 22$$

Es calcula l'entropia:

$$S_{biomarcador} = - \left(\frac{increased}{total\ evidències} \log \frac{increased}{total\ evidències} + \frac{decreased}{total\ evidències} \log \frac{decreased}{total\ evidències} + \frac{no\ change}{total\ evidències} \log \frac{no\ change}{total\ evidències} \right)$$

$$S_{biomarcador} = - \left(\frac{8}{22} \log \frac{8}{22} + \frac{6}{22} \log \frac{6}{22} + \frac{8}{22} \log \frac{8}{22} \right) = 0.473$$

Aquest càlcul s'aplica programàticament a tots el biomarcadors relacionats amb el PTSD, automatitzant la construcció l'URL i la consulta a través de l'API de PTSDDDB.

S'ha considerat filtrar els resultats per aquells biomarcadors que suportessin suficients evidències ($\Sigma\ evidències$) per ser considerats vàlids per l'anàlisi, és a dir, que haguessin estat examinats en un nombre suficient d'estudis d'entre les publicacions contingudes a PTSDDDB. Si no s'apliqués aquesta restricció, biomarcadors amb una sola evidència recollida mostrarien un nivell d'entropia nul, donant una falsa mesura de consens, ja que no hi ha prou indicis per afirmar que no hi ha evidències contradictòries. Donat aquesta premisa es podria deduir que només es poden comparar biomarcadors amb el mateix nombre d'evidències, no obstant es considera que superades $\Sigma\ evidències \Rightarrow 2$ es vàlid fer l'assumpció.

En el següent gràfic es representa la distribució del càlcul d'entropia per a tots el marcadors, mostrant la freqüència de biomarcadors (aquells suportats per al menys $\Sigma evidències \Rightarrow 2$) per un valor determinat d'entropia.

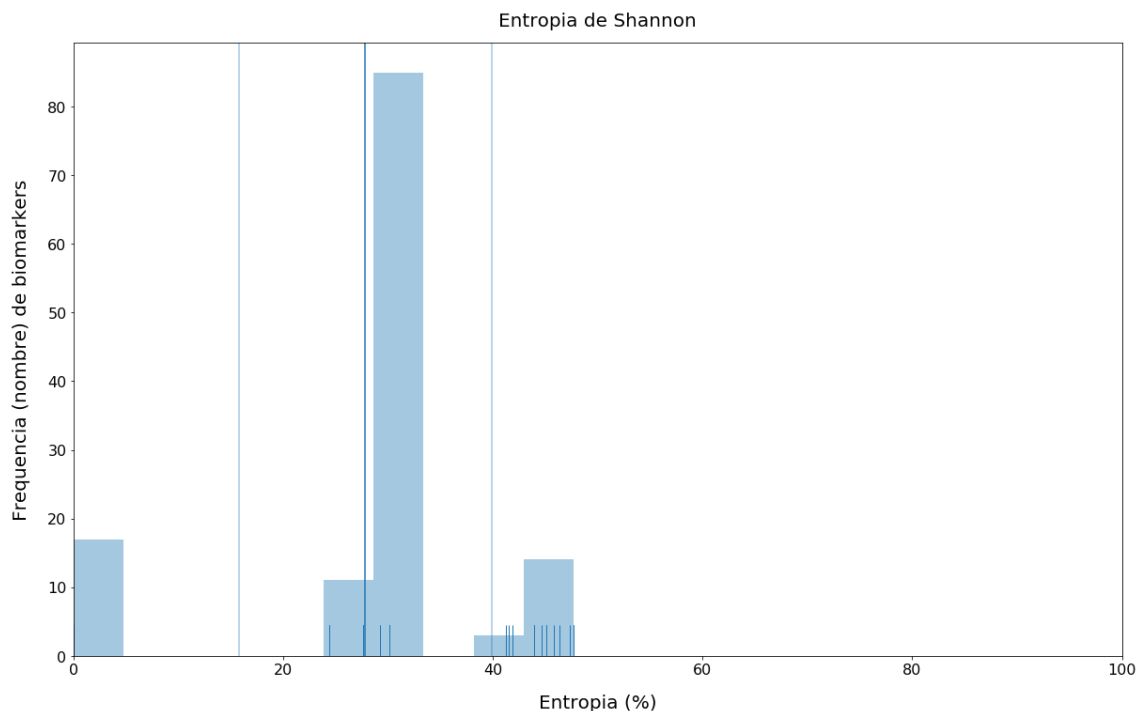


Figura 15. Distribució estadística de l'entropia del canvi relatiu de biomarcadors.

Tenint en compte que l'entropia proporciona una mesura de la discrepància (desordre de les evidències/informació), nivells mínims d'entropia impliquen acord, pel qual el gràfic denota com hi ha consens en un nombre reduït de biomarcadors (18 en concret), mentre que la majoria d'articles mostren un nivell d'evidències contradictòries elevat.

5.2.3.2 Distribució del canvi relatiu dels biomarcadors amb l'entropia

Seguint amb l'anàlisi del consens, s'ha plantejat el càlcul d'una mètrica amb el propòsit d'avaluar la direcció consensuada del canvi relatiu per a cada biomarcador. D'aquesta manera es proposa una ponderació que combini el nombre d'evidències per cada direcció de canvi i penalitzi amb el valor de l'entropia. A continuació es formula aquesta aproximació:

$$direcció\ de\ canvi\ relatiu_{biomarcador} = f_{increased} - f_{decreased} = \frac{j\ evidències\ increased}{total\ evidències} - \frac{j\ evidències\ decreased}{total\ evidències}$$

$$direcció\ de\ canvi\ relatiu\ consensuat_{biomarcador} = direcció\ de\ canvi\ relatiu_{biomarcador} - S_{biomarcador}$$

En el següent diagrama de dispersió es mostra la direcció del canvi relatiu en funció de l'entropia per als diferents biomarcadors de PTSDDDB amb $\Sigma \text{evidències} \Rightarrow 3$. A més, la mida de cada punt expressa la dimensionalitat total de les evidències, el qual és important tenir en compte, ja que es segueix l'assumpció d'aplicar la mateixa mètrica a biomarcadors amb nombre d'evidències diferents.

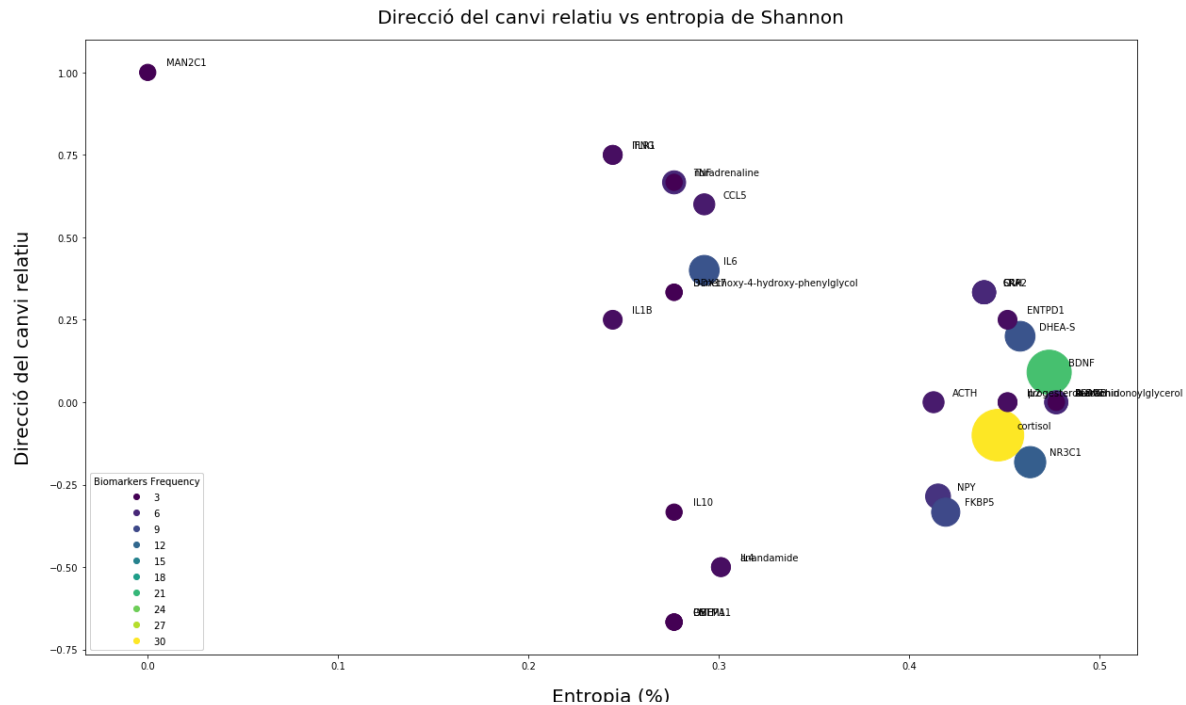


Figura 16. Dispersió estadística de l'entropia del canvi relatiu de biomarcadors.

Tenint en compte la dimensió de la direcció del canvi, es prioritzen aquells biomarcadors que s'allunyen de l'eix de les ordenades (efecte nul), mostrant algun canvi representatiu (incrementat o decrementat) en presència de PTSD. Aquesta consideració respon a la premisa de què per associar un biomarcador a l'efecte d'una enfermetat, s'han de demostrar desregulacions en la presència d'aquests comparant-los amb estudis de control. Per altra banda, seguint la línia de prendre l'entropia com a mesura de consens, es pretén aproximar a l'eix 0 de les absisses.

Es poden anotar a partir del gràfic les següents conclusions:

- La distribució mostra correlació en que a nivells alts d'entropia, la direcció del canvi tendeix a ser nul·la, mentre que a nivells baixos, mostra una direccionalitat més marcada.
- Els biomarcadors que disposen d'àmplies evidències, cortisol i BDNF, 30 i 27 articles respectivament, mantenen nivells d'entropia alts i la direcció del canvi tendeix a aproximar-se a nul·la, pel qual s'afirma que la recopilació d'evidències mostren un consens contradictori.
- Només tendeix a entropia nul·la i amb una direcció de canvi positiva MAN2C1.
- Mantenen una entropia mitjana biomarcadors amb relativament poques evidències (3 -12) i mostren cert nivell de direccionalitat.

Conclusions

En general, aquest projecte representa un primer intent d'integrar el coneixement expressat en dades de resultats d'estudis clínics entorn la psicopatologia del PTSD, amb un enfoc especial en biomarcadors. Actualment, la quantitat creixent de dades, informació i coneixements fa que sigui complicat accedir i tenir en compte tota la informació derivada de la recerca en una àrea d'interès determinada. La versió actual del PTSDDDB, prova com basant-se en la una compilació i organització dels d'estudis publicats i seleccionats i mitjançant un procediment d'extracció acurat de dades, la integració en un patró comú i la visualització interactiva, es facilita a l'investigador una plataforma accessible articulant gran part de la informació relativa a l'àrea d'una patologia.

Pel que fa al resultat final de la plataforma web, es pot afirmar que des de l'esforç d'organitzar les dades derivades de l'extracció d'informació dels articles, s'ha aconseguit elaborar una aplicació web amb planes que presenten el contingut d'una forma ordenada i intuïtiva, proporcionant múltiples funcionalitats d'exploració i portabilitat a l'usuari.

Per altra banda, en referència als anàlisis realitzats sobre les dades obtingudes de les consulta mitjançant l'API de PTSDDDB, es desprèn la manca de consens en la literatura actual entorn l'efecte de la psicopatologia en l'expressió dels biomarcadors putatius.

Donat la quantitat i el valor de dades recollides i organitzades, el projecte presenta molts punts potencials per ampliar, millorar i examinar. Les propostes de desenvolupament futur del projecte s'emmarquen en el propòsit de vincular les metadates d'estudis clínics a les dades d'evidència, obrint la possibilitat de dissenyar models predictius o anàlisis de dades longitudinals. A més, es pretén afegir al projecte models de rutes metabòliques, permetent contextualitzar les metadades en processos biològics a partir de l'enriquiment de rutes basat en biomarcadors.

Reflexió d'autoavaluació i agraïments

El resultat final del projecte que s'exposa en aquesta memòria, és fruit de l'esforç col·lectiu d'un grup d'investigadors. A nivell personal, va suposar per mi el primer projecte en una línia de recerca en el qual se m'assignaven grans parts del desenvolupament, aprenent des de l'estructura d'un estudi clínic, passant per la modelització d'una complexa base dades, fins la programació tant des del *back-end*, polint les meves habilitats de tractament de dades, com des del *front-end*, aprenent a fer ús de tecnologies amb gran potencial de desenvolupament web.

En perspectiva, és satisfactori veure com a partir d'un full de càlcul amb un grau de complexitat elevat, se n'ha derivat una plataforma web amb el contingut organitzat. Gràcies a treballar colze a colze amb *Daniel Domingo-Fernández*, vam aconseguir donar-li forma al projecte cap a un producte final amb usabilitat.

Malgrat que en els propòsits inicials per aquest projecte de final de grau es plantejaven realitzar anàlisis de dades més extensos, la qual cosa es va veure truncada tant per motius de disponibilitat temporal com per privacitat de dades, els anàlisis de consens mitjançant entropia han proporcionat un punt de partida per la continuïtat del projecte, validant la possibilitat de realitzar fàcilment consultes dirigides al contingut organitzat de la base de dades. Avaluant els resultats d'aquest anàlisis de consens cal dir que la integració de quantitats més elevades d'evidències porta contradiccions en el canvi relatiu. Cal destacar que, a caire personal, m'he mostrat escèptic entorn la precisió de les premises establertes entorn als processos biològics de la conducta humana, per lo qual no em van sorprendre els resultats de l'anàlisi de consens.

En primer lloc voldria donar les gràcies a *Daniel Domingo-Fernández* per endinsar-me en la recerca bioinformàtica i transmetre'm la il·lusió en el camp que vaig triar seguir el meu transcurs acadèmic i desenvolupar el meu futur professional. A *Charles Tapley-Hoyt*, per transmetre'm les millors pràctiques de programació i la seva visió audàs i brillant de la bioinformàtica.

A *Alpha Tom-Kodamullil*, per la confiança dipositada en el desenvolupament del projecte PTSDDDB i per transmetre'm la necessitat d'establir contactes i com enfocar i gestionar un projecte complex en l'àmbit de la bioinformàtica amb diversos col·laboradors i participants.

Agrair la confiança del doctor i cap del departament *Martin Hoffmann-Apitius*, per oferir-me l'oportunitat i el suport de formar part del grup de bioinformàtica de l'Institut Fraunhofer

Donar les gràcies a la meva tutora de pràctiques, la doctora *Marta Sales*, per accedir a tutoritzar el treball i dur a terme les discussions i revisions pertinents. També a *Ricardo Cordero* per assessorar-me en els tràmits burocràtics en l'elaboració de la memòria.

Llistat de referències

- [1] Domingo-Fernandez, D., Provost, A., **Marin-Llaó, J.**, Tom-Kodamullil, A., Lasseter, H., Diaz, K., Daskalakis, N., ... & Haas, M. (2019). PTSD Biomarker Database: Deep Dive Meta-database for PTSD Biomarkers, Visualizations, and Analysis Tools. *BioRxiv*, 547901.
- [2] Mayo clinic, informació estrès posttraumàtic:
<https://www.mayoclinic.org/diseases-conditions/post-traumatic-stress-disorder/symptoms-causes/syc-20355967> [consultat 15-05-19]
- [3] Hospital Vall Hebron, informació estrès post-traumàtic:
<https://hospital.vallhebron.com/ca/malalties/trastorn-estres-posttraumatic-tept>
[consultat 15-05-19]
- [4] American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*.
- [5] Mayo clinic, informació tractament PTSD:
<https://www.mayoclinic.org/es-es/diseases-conditions/post-traumatic-stress-disorder/diagnosis-treatment/drc-20355973> [consultat 03-06-19]
- [6] Michopoulos, V., Powers, A., Gillespie, C. F., Ressler, K. J., & Jovanovic, T. (2017). Inflammation in fear-and anxiety-based disorders: PTSD, GAD, and beyond. *Neuropsychopharmacology*, 42(1), 254..
- [7] Informació marcadors clínics
http://www.iispv.cat/recerca/arees_de_recerca/64/grup-de-recerca-en-biomarcadors-de-malalties-i-mecanismes-moleculars-dibiomec-group [consultat 03-06-19]
- [8] Concepte de biomarcador
<http://www.ehu.eus/europeanclass2003/miren1.pdf> [consultat 03-06-19]
- [9] McEwen, B. S. (2008). Central effects of stress hormones in health and disease: Understanding the protective and damaging effects of stress and stress mediators. *European journal of pharmacology*, 583(2-3), 174-185.
- [10] Galvez-Contreras, A. Y., Campos-Ordonez, T., Lopez-Virgen, V., Gomez-Plascencia, J., Ramos-Zuniga, R., & Gonzalez-Perez, O. (2016). Growth factors as clinical biomarkers of prognosis and diagnosis in psychiatric disorders. *Cytokine & growth factor reviews*, 32, 85-96.
- [11] IL6
<https://www.ncbi.nlm.nih.gov/gene/3569> [consultat 03-06-19]
- [12] Binder, E. B. (2009). The role of FKBP5, a co-chaperone of the glucocorticoid receptor in the pathogenesis and therapy of affective and anxiety disorders. *Psychoneuroendocrinology*, 34, S186-S195.

[13] NR3C1

<https://www.genecards.org/cgi-bin/carddisp.pl?gene=NR3C1> [consultat 27-05-19]

[14] DHEAS

<https://www.mayoclinic.org/es-es/drugs-supplements-dhea/art-20364199> [consultat 27-05-19]

[15] PTSDpubs

https://www.ptsd.va.gov/ptsdpubs/search_ptsdpubs.asp [consultat 27-05-19]

[16] Glossari d'assajos clínics

<https://www.nia.nih.gov/research/dgcbg/clinical-research-study-investigators-toolbox/nia-glossary-clinical-research-terms> [consultat 27-05-19]

[17] Django

<https://developer.mozilla.org/es/docs/Learn/Server-side/Django/Introducci%C3%B3n>

[consultat 03-06-19]

[18] APIRest

<https://bbvaopen4u.com/es/actualidad/api-rest-que-es-y-cuales-son-sus-ventajas-en-el-desarrollo-de-proyectos> [consultat 03-06-19]

[19] Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, 27(3), 379-423.

Annex 1

*Normativa de Treball Fi de Grau Facultat d'Enologia
Aprovada per Junta de Facultat d'Enologia del dia 30 d'octubre de 2014*

ANNEX 1

FORMULARI SOL·LICITUD DEL TREBALL DE FI DE GRAU

Alumne: Josep Marin Llaó

Correu-e : josep.marin@estudiants.urv.cat

Títol provisional *PTSD Knowledge Base: Meta-database confection and analysis for PTSD Biomarkers*

Resum del treball

La base de dades de biomarcadors de PTSD (PTSDDDB, desenvolupada en col·laboració durant l'estada de practiques que vais realitzar al Fraunhofer Institut) proporciona un overview dels biomarcadors fisiològics que s'estan estudiant com a marcadors putatius en la literatura actual de trastorns d'estrès posttraumàtic (PTSD o TEPT) per permetre als investigadors explorar i comparar els resultats. PTSDDDB conté actualment més de 900 biomarcadors, i la seva informació rellevant, extrets de 109 articles publicats entre 1997 i 2017. El contingut emmagatzemat en aquesta base de dades es representa amb una aplicació web que ofereix una interfície amb múltiples visualitzacions interactives que permeten investigar el coneixement del biomarcador en el TEPT/PTSD .

A banda d'algunes tasques menors de desenvolupament web pendents al projecte i les que sorgeixin de noves visualitzacions a partir de nous enfocaments d'anàlisi o d'importació de dades, el treball durant els pròxims mesos a realitzar com a TFG consistiria fonamentalment en confecció de la memòria recollint tot el realitzant en el projecte i realitzar de manera més acurada l'anàlisi de dades, amb diferents mètodes com poden ser l'anàlisi multivariant amb PCR, i concretar les conclusions biològiques i/o clíniques que se'n derivin.

Signatura del Alumne/a



Tarragona a 30 de Gener de 2019

Annex 2

*Normativa de Treball Fi de Grau Facultat d'Enologia
Aprovada per Junta de Facultat d'Enologia del dia 30 d'octubre de 2014*

ANNEX 2

FITXA DE SEGUIMENT DEL TUTOR/A del TFG

Nom i Cognoms de l'Alumne/a: _____ Josep Marín Llaó _____

Nom i Cognoms del Tutor/a: _____ Marta Sales Pardo _____

Data de la entrevista amb l'alumne: __ 15/03/2019, 03/06/2019, 06/06/2019 _____

Recomanacions durant el seguiment: __ Parlar amb els responsables de a base de dades sobre la possibilitat d'utilitzar datasets concrets per predir el risc de desenvolupament de PTSD.

Això no va ser possible, i per tant es va optar per fer una anàlisi dels biomarcadors putatius d'acord amb les dades públiques a través de la base de dades.

Observacions: __ El Josep treballa de manera totalment autònoma i ha sabut trobar sortides al projecte quan no ha estat possible dur a terme el pla original. __

Observacions Darrera revisió:

_ Hem acordat afegir una anàlisi dels processos relacionats amb els biomarcadors de PTSD més rellevants utilitzant les eines de difusió que ell mateix ha realitzat en un projecte de recerca a la UPC.

Signatura del Tutor/a



Signatura del Alumne/a

Tarragona a __ 6 __ de __ juny __ 2019 __

[illegible]

Annex 4

PTSDDDB_API_TUTORIAL

June 17, 2019

1 PTSDDDB Relative Changes

1.0.1 Investigate the relative changes of the biomarkers in the database

Author: Josep Marin-Llaó

More details about the PTSDDDB API in the links below: - Swagger UI: <https://ptsd.scai.fraunhofer.de/swagger-ui> - Database Schema: https://ptsd.scai.fraunhofer.de/static/ptsd/img/db_schema.png
Python libraries required

```
In [1]: import time
import sys
import json
import urllib.request
from collections import defaultdict

import numpy as np
import matplotlib.pyplot as plt
from tqdm import tqdm
```

```
In [2]: time.ctime()
```

```
Out[2]: 'Sat Jun  8 20:41:44 2019'
```

```
In [3]: sys.version
```

```
Out[3]: '3.7.3 (default, Mar 27 2019, 09:23:15) \n[Clang 10.0.1 (clang-1001.0.46.3)]'
```

```
In [4]: %matplotlib inline
```

Load data from API endpoint. This endpoint returns all the relative changes for all the biomarkers in the database as a JSON object

```
In [5]: with urllib.request.urlopen("https://ptsd.scai.fraunhofer.de/api/biomarker/all") as url:
all_biomarkers = json.loads(url.read().decode())
```

```
In [6]: def get_effect_biomarker(biomarker):
url_biomarker = "https://ptsd.scai.fraunhofer.de/api/biomarker_effect/"+biomarker
with urllib.request.urlopen(url_biomarker) as url:
return(json.loads(url.read().decode()))
```

