
This is the **published version** of the bachelor thesis:

Pol Esteve, Laura; Martí Godia, Enric, dir. Sistema d'integració, visualització i anàlisi de dades genètiques pel seguiment de malalties minoritàries. 2022. (1394 Enginyeria de Dades)

This version is available at <https://ddd.uab.cat/record/264634>

under the terms of the  license

Sistema d'integració, visualització i anàlisi de dades genètiques pel seguiment de malalties minoritàries

Laura Pol Esteve

Resum—La identificació de canvis en la seqüència o l'estructura de l'ADN pot conduir a determinar la causa que hi ha darrera de l'aparició d'una malaltia genètica. Diferents tècniques d'anàlisi de l'ADN han esdevingut rutinàries en la majoria de centres hospitalaris i centres de recerca per a l'estudi simultani dels gens que conté el genoma humà. Aquest tipus de tecnologia s'anomena seqüenciació massiva o Seqüenciació de Nova Generació (NGS) i s'utilitza habitualment en l'àmbit de les malalties minoritàries junt amb la tècnica d'array d'Hibridació Genòmica Comparativa (a-CGH). L'objectiu d'aquest projecte és integrar els resultats d'aquestes diverses tecnologies tot creant una plataforma on es puguin volcar, estructurar, visualitzar i analitzar les dades genètiques del Servei de Medicina Genètica i Molecular de l'Hospital Sant Joan de Déu per tal de poder fer un seguiment de les diferents patologies identificades per departament i fenotip, aconseguint així millorar i accelerar el diagnòstic genètic de pacients amb malalties rares. Així mateix, mitjançant la plataforma es podran caracteritzar les variants detectades en els pacients que s'hagin estudiat amb ambdues tecnologies fent comparatives de sensibilitat i especificitat.

Paraules clau—ADN, gen, fenotip, a-CGH, NGS, CNV, SNV, mutació, deleció, duplicació, variant genètica, prova genètica, malaltia genètica, base de dades, django, mongodb, wireframe, website.

Abstract—The identification of changes in the sequence or structure of DNA can lead to determining the cause behind the occurrence of a genetic disease. Different DNA analysis techniques have become routine in most hospitals and research centers for the simultaneous study of the genes contained in the human genome. This type of technology is called massive sequencing or Next Generation Sequencing (NGS) and is commonly used in the field of rare diseases together with the array Comparative Genomic Hybridisation (a-CGH) technique. The aim of this project is to integrate the results of these different technologies by creating a platform where the genetic data of the Genetic and Molecular Medicine Service of the Hospital Sant Joan de Déu can be dumped, structured, visualized and analyzed in order to monitor the different pathologies identified by department and phenotype, thus improving and accelerating the genetic diagnosis of patients with rare diseases. In addition, the platform will allow the variants detected in patients studied with both technologies to be characterized, making comparisons of sensitivity and specificity.

Index Terms—DNA, gene, phenotype, a-CGH, NGS, CNV, SNV, mutation, deletion, duplication, genetic variant, genetic test, genetic disease, database, django, mongodb, wireframe, website.

1 INTRODUCCIÓ

Segons l'Organització Mundial de la Salut un registre de pacients és un fitxer de documents que conté informació uniforme sobre persones individuals, recollida de forma sistemàtica i integral, per què pugui ser utilitzada per respondre a objectius establerts prèviament, ja siguin científics, clínics o de política sanitària [1].

Els registres de pacients són una eina fonamental per a la investigació biomèdica (ja sigui clínica, bioquímica, genètica, de la correlació entre fenotip/genotip, del diagnòstic, tractament o evolució, entre d'altres), i per a la investigació epidemiològica en general.

Aproximadament el 80% de les malalties minoritàries tenen una base genètica, i identificar els gens que provoquen canvis (mutacions) en la malaltia té molts

beneficis per als pacients i les famílies:

- Obtenir un diagnòstic precís.
- En alguns casos, un diagnòstic genètic precís i precoç condueix a teràpies específiques dirigides que poden tractar de manera més eficaç la malaltia o, en casos rars, fins i tot poden revertir o prevenir la malaltia.
- El diagnòstic genètic es pot fer extensible en altres familiars i definir els riscos de recurrència, així com oferir el corresponent assessorament genètic.
- Entendre els gens i les proteïnes associats a les malalties és fonamental per avançar en la comprensió de les malalties genètiques.

Malaauradament, el diagnòstic genètic és complex i sovint requereix temps, cosa que pot afectar negativament als pacients i les famílies. Es calcula que només el 40% de les

proves acaben en un diagnòstic definitiu. Com a resultat, un gran nombre de proves acaben amb resultats negatius o no concloents perquè no hi ha prou evidència que els canvis identificats siguin la causa de la malaltia en els pacients estudiats.

En el cas de les malalties minoritàries (rars) els registres són especialment importants a causa de l'escàs nombre d'afectats per cadascuna. Per aquest motiu, el fet de poder accedir a un registre a través d'una plataforma que pugui volcar, estructurar, visualitzar i analitzar aquestes dades augmenta exponencialment la capacitat de gestionar els pacients amb malalties genètiques. Facilita no només l'agrupació i emmagatzemament de les dades genètiques dels pacients en una única estructura, sinó també que es comparteixin les dades entre diferents membres del departament, evitant redundància i millorant l'organització.

La idea de crear una plataforma per emmagatzemar les dades genètiques dels pacients va ser proposada pel departament de Genètica de l'Hospital Sant Joan de Déu, ja que actualment no tenen un sistema definit per poder realitzar aquest procés de sincronització de les dades. Tot i no estar implementat, tinc constància que altres hospitals disposen d'una eina semblant, però degut a que les dades no són d'ús públic, no s'han trobat referents semblants.

Cal mencionar que existeix el *Beacon Network*[14], un motor de cerca global de mutacions genètiques compartit entre diferents organitzacions. A més a més, a principis del 2020, la Fundació de La Marató de TV3 va iniciar un projecte per implementar una "Xarxa Interhospitalària Catalana de Variants Genètiques per Millorar el Diagnòstic Genètic en Malalties Rares" en el qual s'hi inclou l'Hospital Sant Joan de Déu com a coordinador. Aquest projecte s'està duent a terme a dia d'avui i s'estan definint les ontologies comunes de totes les bases de dades dels hospitals de Catalunya que hi participen (tretze actualment).

Fins ara, han estat emmagatzemant les dades dels pacients a través d'excels modificats manualment. A causa de problemes amb la coherència i poca gestió amb les dades, han decidit fer un plantejament global i proposar crear una base de dades que es sincronitzi amb una website perquè puguin anar afegint dinàmicament les dades i així no haver de fer-ho manualment.

Per tal de guiar-me i poder realitzar un disseny de la plataforma, s'han buscat diverses plantilles de websites de diferents hospitals.

Tenint en compte això, l'objectiu principal del projecte consisteix en dissenyar i implementar una website on poder volcar, estructurar, visualitzar i analitzar les dades genètiques dels pacients en forma de variant o canvis genètics.

A més a més, es pretén assolir funcionalitats com:

- La integració de resultats de diverses tècniques d'anàlisi de l'ADN en una base de dades.
- La visualització de mètriques de pacients.
- L'anàlisi d'aquestes dades per tal d'extreure

conclusions i poder fer una comparativa entre diferents tècniques.

En la creació d'aquesta plataforma s'integraran les dades de les tècniques d'array d'Hibridació Genòmica Comparativa (*a*-CGH – *Comparative Genomic Hybridization Array*)[17] i de Seqüenciació de Nova Generació (NGS – *Next Generation Sequencing*)[18], i es crearà una website on s'introduiran els resultats d'aquestes dues tècniques on constaran les variants informades amb els criteris de classificació estandarditzats.

Els principals beneficis de la creació de la plataforma són:

- Establir una estructura de base de dades comuna que pugui servir a altres departaments on es realitzin proves genètiques perquè en un futur es pugui interrogar de manera eficient diferents mutacions entre diferents serveis i/o institucions.
- Organitzar informació complexa sobre fenotips utilitzant termes estàndard comuns que permetin la comparació i interpretació de les dades.
- Analitzar i comparar les diferents tècniques que s'utilitzen per detectar alteracions a l'ADN, incloses microdeleccions i microduplicacions

L'eina de desenvolupament *Balsamiq*[2], s'utilitzarà per a dissenyar el frontend de l'aplicació per tal de poder visualitzar i estructurar la interacció que tindran els usuaris (personal de cada departament) amb l'aplicació, i segons la secció poder veure on s'introduiran els inputs (variants). A més a més, a través de l'eina *Moon Modeler*[3], es crearà un disseny d'esquema per la base de dades MongoDB on es podrà veure l'estructuració d'aquestes dades amb les seves col·leccions, documents, variables i connexions entre aquestes.

Sabent que el departament de genètica té un gran flux de dades, és ideal que la base de dades pugui ser escalable sota demanda. Per aquest mateix motiu, s'ha escollit utilitzar una base de dades no-relacional, com *MongoDB*[4]. No només permet un modelatge de les dades gràcies al seu emmagatzematge a través d'estructures JSON, sinó que també proporciona una alta disponibilitat, i al ser *Schema Less* i basada en documents, ens permet tenir flexibilitat.

Per tal d'integrar les dades amb l'aplicació, s'ha decidit utilitzar diferents eines. Per una banda *Django Rest Framework* *CRUD Rest API*[5] i per l'altra, utilitzaré *Djongo*[6] juntament amb *MongoEngine*[7] per a integrar el frontend amb el backend.

A més a més, amb l'objectiu d'accedir més fàcilment a la base de dades, s'utilitzarà *MongoDB Compass*[8], eina interactiva la qual permet consultar, optimitzar i analitzar les dades de MongoDB.

Amb la finalitat de visualitzar les dades, utilitzaré eines com *Tableau*[9] i *R Shiny proxy*[10].

Tot el codi que es vagi implementant durant el desenvolupament del projecte, s'anirà penjant en un repositori de *GitHub*[11] per a poder fer un seguiment constant i tenir un control de les modificacions.

2 REQUERIMENTS

Per tal de familiaritzar-me amb el format de les dades de l'hospital i conèixer els requeriments, s'han realitzat un seguit de reunions amb el Dr. Guerau Fernandez Isern, facultatiu del Servei de Medicina Genètica i Molecular de l'Hospital Sant Joan de Déu i persona responsable de comunicar les necessitats del departament de genètica, i en Joe Kaen, facultatiu de IT systems.

El TFG anirà enfocat exclusivament al departament de Genètica. És cert que hi ha dos serveis més (Hematologia i Oncologia Pediàtrica) amb qui també estem en el procés de definir les variables a guardar, però ara per ara el més avançat és el de Genètica.

A continuació es destaquen els requeriments extrets en aquestes reunions. Per una comprensió clara dels requeriments establerts, s'han dividit en requeriments funcionals i requeriments de dades.

2.1 Requeriments funcionals

Dins del servei de Genètica es poden realitzar mostres a pacients a través de dues metodologies: NGS i a-CGH. Per tant, els inputs que rebrem estan dividits en dues seccions. Per una banda les dades de realitzar una mostra amb la tècnica NGS i per altre banda, les dades resultants de la tècnica d'array a-CGH.

La principal idea és que cada servei tingui la seva base de dades pròpia, i dins de la base de dades s'emmagatzemin els resultats d'un pacient per a diferents mostres, ja siguin d'una tècnica i/o una altre. Les dades proporcionades per part del departament de Genètica estan agrupades en variables per tres categories: variants totals, variants informades i fenotip.

A continuació s'explica el circuit d'entrada de dades (inputs) i el circuit de sortida de dades (outputs)

a) Inputs

Per la base de dades del departament de genètica, tindrem diferents fitxers amb informació sobre les dues tècniques que s'utilitzen, NGS i a-CGH.

Ambdues tècniques són eines genòmiques per a la detecció de CNV (*Copy Number Variant*)[12]. A més a més, NGS té la capacitat de detectar SNV (*Single Nucleotide Variant*)[13]. Cada una d'aquestes tècniques tindrà diferents camps d'entrada per a ser consultats:

- Dades genètiques: dades genòmiques resultants de realitzar mostres a un pacient.
 - Dades variants totals: dades variants provinents de totes les variants trobades als pacients. Hi poden haver dades CNV i SNV.
 - Dades variants informades: són dades informades reportades a l'informe genètic, és a dir, dades provinents de les variants totals que s'han considerat significatives per la patologia dels pacients en estudi.
- Dades fenotípiques: són dades provinents de les característiques observables dels pacients.
- Dades tècniques: dades provinents de la

metodologia emprada en el procés de seqüenciació. Són dades comunes tant per variants totals com per variants informades.

Per a la tècnica NGS rebrem quatre fitxers en format tsv corresponents a: dades tècniques, dades fenotípiques, variants totals de les CNV i variants totals de les SNV. Les variants informades seran introduïdes a través d'un formulari pel frontend.

Per altre banda, per la tècnica a-CGH, tot i estar encara pendent de concretar amb el departament, pel desenvolupament del projecte s'ha considerat que rebrem el mateix format de dades que per la tècnica NGS.

b) Outputs

Pel que fa al circuit de sortida de les dades (outputs), tindrem una visualització externa i interna.

- La visualització externa fa referència a les dades agregades a través del *Beacon Network*[14]. Aquestes dades podrien ser interrogades per altres hospitals i es composaria d'un subset de totes les variables que recopila el departament de genètica.
- La visualització interna es podria dividir en dos aspectes.
 - 1) Visualització de dades estadístiques (mostres per servei, número de mostres d'una tècnica per nom de la mostra, pacients informats amb mutació causant de la patologia identificada, solapament entre CNVs de NGS i arrays, etc)
 - 2) Visualització de dades per pacient, on indicant el "NHC - Numero d'Història Clínica" puguem veure les variants, o consultant per regió cromosòmica puguem veure quants pacients tenen certa mutació, sigui SNV o CNV. En aquest cas, les dades serien per consulta interna i per tant es podrien indicar les dades identificatives dels pacients.

A més a més, tot i mostrar una visualització de les dades amb informació rellevant pels doctors del departament de Genètica, es realitzarà un anàlisi de les dades per comprovar diferents hipòtesis plantejades.

2.2 Requeriments de dades

S'hauran de tenir en compte un seguit de requisits per tal de dissenyar la base dades:

- Cada pacient s'identifica amb el Número d'Història Clínica (NHC), el SampleType (excepció pel cas de tenir dos individus amb el mateix NHC, és a dir, per identificar el fetus i la mare) i el PetiDate (per identificar mostres d'un mateix pacient al llarg del temps en el cas d'oncologia).
- Un pacient pot haver-se realitzat més d'un tipus de tècnica de seqüenciació (ja sigui NGS o a-CGH).
- Cada tècnica de seqüenciació (NGS i a-CGH) s'identifica amb PoolNum i SampleName.
- Cada tècnica pot tenir més d'una variant total.
- Cada tècnica pot tenir més d'una variant informada.

- Les dades fenotípiques són comunes per tècnica, és a dir, un pacient tindrà les mateixes dades fenotípiques per cada tècnica, sigui NGS o a-CGH
- Les variants totals tenen com a camps comuns les dades tècniques de la tècnica utilitzada (NGS o a-CGH).
- Igual que les variants totals, les variants informades tenen exactament les mateixes dades tècniques.
- Per a cada variant hi pot haver dades CNV o dades SNV, és a dir, una variant pot ser SNV o CNV, però no pot ser ambdues alhora.
- Algunes dades fenotípiques del pacient poden variar al llarg del temps. Per això en comptes de modificar aquestes dades, es guardarà un històric.
- El camp Omic indica el tipus de tècnica de seqüenciació. Si Omic és "DNA-seq" significa que la tècnica correspon a NGS, i si Omic és "CGH-array" voldrà dir que és la tècnica a-CGH.
- Poden haver-hi dades "None", que no existeix el camp, o dades "NaN", dades que no tenen informació.

Adicionalment, s'han de contemplar els següents requisits,

- **Redundància:** Es pot produir la presència del mateix identificador per dues mostres diferents. Aquest esdeveniment és a causa de la identificació de mostres provinents del fetus matern ja que se l'identifica amb el NHC (Número d'Història Clínica) de la mare. Les variants totals tenen una variable anomenada SampleType que indica el tipus de mostra (de quin teixit prové). S'haurà d'identificar amb el tipus de teixit: si és teixit en sang serà la mare, i si és teixit fetal serà el fetus.
- **Temporalitat:** Una mateix pacient pot tenir dues entrades utilitzant la mateixa tècnica de detecció. Això pot produir-se a partir d'un augment en la quantitat de gens analitzats (panell de gens vs tots els gens) o si es tracta d'un pacient oncològic pot determinar diferents estadis de la patologia. En aquest segon cas les variants hauran d'estar etiquetades amb la data de petició de la prova (PetiDate) per poder-les diferenciar.
- Les dades CNV tindran com a posició del cromosoma les variables "start" i "end" amb valors diferents. En canvi les dades SNV tindran el mateix valor tant per "start" com per "end". Inicialment estava guardat com una sola variable anomenada posició, però per tenir-les igual, s'ha afegit start i end per les dues.

3 DISSENY

En aquest apartat es defineix el disseny de la website, començant pel disseny de la base de dades (backend) i acabant pel disseny de la pròpia interfície (frontend).

3.1 Disseny de la Base de Dades

Després de dissenyar diverses versions segons les

necessitats del departament de Genètica, pel disseny final de la base de dades s'han dividit els camps entre específics i comuns per tal d'evitar emmagatzemar dades duplicades. Els camps específics engloben les variables exclusives per cada arxiu de SNV i CNV, i els camps comuns inclouen, per una banda les variables que es poden extreure comunes de les variants SNV i CNV, i per l'altra, engloba els camps únics per cada pacient (que serien les dades fenotípiques), i els camps comuns per a cada tècnica.

A més a més, s'han detectat diferents equivalències entre fitxers amb variants reportades, com serien "DATA NAIX i DOB", "TECNICA i Omic" o "SERVEI i Servei" entre d'altres.

Per tant, en la versió final s'ha creat una base de dades on s'hi han inserit dues col·leccions (Figura 1):

- La primera col·lecció, anomenada "**Patients**", conté informació per cada pacient incloent les dades fenotípiques que no variaran (emmagatzemades com un objecte) i les dades fenotípiques que poden variar (emmagatzemades com un array per guardar l'històric de les dades anteriors). Cada pacient s'identifica amb un "_id" autogenerat per mongo. Com que un pacient pot haver-se fet diferents proves tant de la tècnica NGS com de la tècnica a-CGH, s'ha creat un array de NGS i un per a-CGH. Cada array conté totes les proves que s'han realitzat de cada tècnica (indicada pel camp "Omic"), i per identificar-les, s'ha definit un "_id" de mongo per així relacionar cada identificador de l'objecte contingut en l'array amb la mostra en si.
- Cada mostra té unes dades genètiques específiques. Per això s'ha decidit emmagatzemar-les en una segona col·lecció anomenada "**Samples**", que conté la informació genètica del pacient segons la tècnica de seqüenciació utilitzada. Com poden haver-hi moltes proves de cada tècnica per un sol pacient, hi ha tants documents com proves hi hagi, i s'identifiquen amb els camps "PoolNum" i "SampleName" (que es relacionen amb els arrays de NGS i a-CGH). Cada mostra pot contenir diferents CNVs i SNVs, camps els quals s'ha decidit emmagatzemar-los en arrays d'objectes, ja que cada CNV i SNV conté diversos camps ("CHROM", "START", "END", "Length", etc).

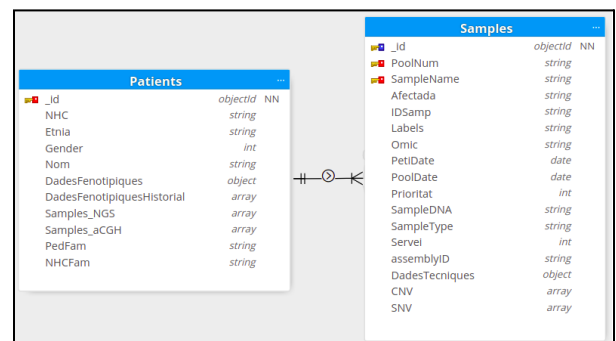


Figura 1. Base de dades del departament de Genètica dissenyada amb l'eina MoonModeler

3.2 Disseny de la interfície

Amb l'objectiu de crear un disseny de la plataforma, s'ha utilitzat l'eina de disseny Balsamiq per crear una plantilla inicial i plasmar visualment els requeriments definits junt amb el departament de genètica.

Aquests requisits inclouen que donat un usuari qualsevol, aquest es pugui logar segons si treballa al departament de Genètica, Oncologia Pediàtrica o Hematologia. Depenent del departament, aquest usuari haurà de tenir uns permisos o altres. Si per exemple és del departament de Genètica, només podrà tenir accés a visualitzar i editar les dades del seu departament i per altra banda podrà accedir a les dades globals de l'hospital com tots els usuaris.

S'han creat diferents dissenys d'interfície (wireframes), un per cada pàgina web linkada a la pàgina principal. Inicialment, quan un usuari accedeix per primer cop, la primera pantalla que visualitzarà serà el "dashboard", i seguidament podrà navegar a través de les diferents seccions del menú i visualitzar les dades genètiques dels pacients o afegir un pacient.

El primer wireframe creat és la pàgina d'inici de sessió "login page" que permetrà als usuaris logar-se amb el seu nom i departament corresponent.

S'ha dissenyat un menú per poder navegar per la pàgina web i així accedir a les diferents funcionalitats. A més a més, s'ha dividit la pàgina en cinc seccions, una enfocada per l'hospital, una altra per cada departament, el de Genètica, Oncologia Pediàtrica i Hematologia, i finalment hi trobem una secció per l'administració de la pàgina.

Dins la secció de l'hospital s'hi troben dues subseccions, una anomenada "dashboard" i una altra anomenada "departments". El "dashboard" (Figura 1 de l'apèndix), és la pàgina que es mostra a la pantalla principal per defecte i proporciona als usuaris una visualització externa on s'hi pot veure una visió general d'estadístiques de l'hospital i visualitzacions de dades d'anàlisi predictiva. Per altra banda, la subsecció "departments" (Figura 2 de l'apèndix) mostra una visió global dels diferents departaments que té l'hospital.

En la secció dels departaments s'hi troben també dues subseccions. La primera seria la que rep el nom de "All patients" (Figura 3 de l'apèndix), que proporciona una visualització de tots els pacients registrats a la base de dades i permet fer una cerca d'aquests i la segona és l'anomenada "Add patient" (Figura 4 de l'apèndix), la qual permet afegir un pacient junt amb les seves mostres realitzades per cada tècnica de seqüenciació (NGS i a-CGH).

4 DESENVOLUPAMENT

La implementació de la plataforma s'ha dividit en dos apartats. El primer es correspon a la implementació de la base de dades amb MongoDB i el segon a la implementació de la website amb Django i MongoEngine. Posteriorment s'explica la visualització i l'anàlisi de les dades.

4.1 Implementació de la base de dades

Com que les dades dels pacients estan emmagatzemades per cada departament en fitxers, primer s'ha hagut de fer un procés de transformació i processat de les dades per tal que no hi haguessin incoherències tot modificant el format de les dades d'entrada dels fitxers i extraient les variables a inserir en cada col·lecció. Una vegada feta aquesta transformació i definides les ontologies i el format de les dades esperat, aquestes s'han inserit a una base de dades no relacional (MongoDB) a través de la creació d'un script en python per convertir els camps en format json.

S'han utilitzat les llibreries de python *pandas*[15], per poder manipular les dades, i *pymongo*[16] per tal de connectar-se des del Pycharm amb la connexió en local de MongoDB.

El script creat, anomenat *InsertData.py*[19], conté una classe global que està inicialitzada amb el nom de les col·leccions i de totes les connexions necessàries. Dins d'aquesta classe hi trobem una funció per comprovar que el dataset passat com a paràmetre existeix a la base de dades. Per altra banda, la classe conté una altra funció que elimina el dataset que li passem com a paràmetre.

Abans d'inserir les dades, es comprova si la base de dades ja existeix, i en el cas de ser així, s'elimina per poder tornar-la a crear.

Troblem una tercera funció que donats uns fitxers com a paràmetres, els obre i processa el contingut per tal de transformar-lo al tipus esperat per Mongo. Una vegada aquestes dades estan transformades, aquesta funció insereix les dades dels fitxers a la col·lecció "Patients" i "Samples" tot relacionant-les amb els camps "PoolNum" i "SampleName" continguts en tots els fitxers.

Per tal d'inserir els pacients, aquests s'han identificat pel Número d'Història Clínica (NHC) i el tipus de teixit (SampleType). En el cas de tenir un pacient amb el NHC repetit i amb el tipus de teixit fetal, indicat amb una "F_" inicial, aquest pacient, tot i tenir el mateix NHC, haurà de ser inserit a la base de dades com un nou document (pacient). En canvi, si tenim una mostra que s'ha realitzat en teixit en sang perifèrica, indicat com "SP", per un pacient ja existent, en aquest pacient significa que se li ha realitzat una altra mostra i per tant s'haurà d'actualitzar l'array de mostres per la tècnica específica. El script crea una col·lecció "Samples", la qual conté les mostres realitzades per cada tècnica a cada pacient (crea un document per mostra) i a la vegada, dins la col·lecció "Patients" afegeix l'identificador de la mostra dins la llista de mostres "Samples_NGS" o "Samples_aCGH".

Per últim, s'han creat dues funcions per inserir dins la col·lecció "Samples" totes les CNV i SNV que hi hagi per cada mostra.

4.2 Implementació de la website

Amb totes les dades introduïdes a la base de dades, la idea inicial era utilitzar enginyeria inversa per tal de revertir les col·leccions en models de Django. Un model és la font única i definitiva d'informació sobre les dades. Conté els camps i comportaments essencials de les dades

que s'estan emmagatzemant. Django, gràcies als models, ofereix una API d'accés a la base de dades generada automàticament. El problema és que Django està originat per treballar amb bases de dades relacionals, i com en aquest cas utilitzem una no relacional com és MongoDB, la enginyeria inversa per generar automàticament el model esperat no és compatible. Per aquest motiu, s'ha hagut de crear manualment el fitxer `models.py` comprovant que els camps i els tipus de dades especificats siguin idèntics a les dades emmagatzemades a la base de dades MongoDB.

Com tota website creada amb Django, cada aplicació conté els fitxers inicials autogenerats, a partir dels quals s'ha anat implementant el codi per la creació de la plataforma[20].

Per configurar i redirigir correctament les pàgines en cada cas, s'ha modificat els fitxers `"urls.py"` tant de l'aplicació com de la carpeta de configuració, i el fitxer `"settings.py"` per tal de connectar el frontend amb la base de dades no relacional creada.

S'ha originat una carpeta anomenada `templates` on s'hi han creat diferents plantilles per cada pàgina redirigida de l'aplicació. Una plantilla és un fitxer de text que defineix l'estructura o la disposició d'un fitxer (com ara una pàgina HTML). Les diferents plantilles (de l'anglès "templates") creades s'han dividit en tres carpetes diferents: "Partials" que conté la base de l'aplicació i la plantilla del menú, "Patients" amb totes les plantilles de visualització dels pacients i de les mostres, i finalment la carpeta "Users" que conté les plantilles de login, logout i perfil de l'usuari que està accedint a la website.

Per tal d'entendre el funcionament de la website, s'ha creat un diagrama de flux (Figura 5 de l'apèndix) que mostra tot el procés desde que un usuari accedeix fins que marxa de la pàgina. Les figures geomètriques representen cada pas del procés i es connecten entre si a través de fletxes i línies que marquen la direcció del flux.

La primera vegada que s'accedeix a la website, automàticament es redirigeix a la pàgina d'accés dels usuaris (Figura 6 de l'apèndix). En aquest cas del departament de genètica, i es demana que s'introdueixi el nom i la contrasenya. Depenent del departament en que treballi l'usuari que està accedint, tindrà uns permisos o uns altres. A la base de dades s'ha creat una col·lecció anomenada `"users_profile"` que emmagatzema tots els perfils que es van creant dels usuaris de cada departament. Si l'usuari és administrador, podrà tenir accés a les bases de dades de tots els serveis. En canvi si l'usuari està per exemple en el grup de genètica, aquest només tindrà accés a la base de dades de genètica.

Logat l'usuari, aquest passa a veure la pàgina d'inici on es mostra una visió global de les dades (Figura 2). A l'esquerra s'hi pot observar el menú de navegació i a la part superior s'hi troba un menú en gris que mostra "Logout" o "Login" depenent de si l'usuari s'ha autenticat o no. Tal com s'havia dissenyat en l'apartat de "Disseny de la interfície", el menú permet a l'usuari navegar i seleccionar si visualitzar el seu perfil, el

dashboard, informació dels departaments i informació dels pacients, afegir un pacient o accedir a la configuració. La pàgina que es mostra a la pantalla d'inici per defecte és el **"Dashboard"**, el qual mostra gràfiques que permeten tenir una visualització estadística dels pacients i de les mostres. Per una banda, a la part superior (Overview), es mostra informació sobre la última actualització de la base de dades i del número de pacients, mostres i usuaris emmagatzemats en les tres col·leccions de la base de dades. I per altre banda, una visualització interna de les mostres realitzades per la tècnica NGS (ja que per a-CGH encara no tenim les dades) i un anàlisi de les dades que s'explicaran en els apartats "Visualització de les dades" i "Anàlisi de les dades".

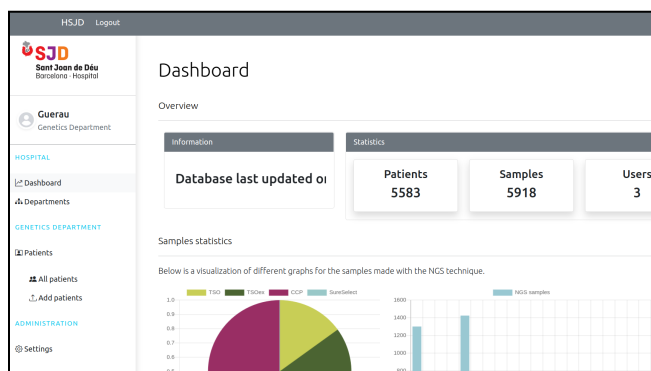


Figura 2. Output del fitxer `dashboard.html`

Navegant a través del menú, es pot accedir al perfil de l'usuari que ha accedit a la pàgina i veure la informació bàsica de la persona (Figura 3). En cas de tenir una imatge del perfil, aquesta apareixerà a la dreta del requadre.

The screenshot shows a 'Profile Information' form. The form fields and their values are: First name (Guerau), Last name (guerau), Username (guerau), Email (example@gmail.com), Phone (685132589), Address (HSJD), and Department (Genetics). To the right of the form is a placeholder for a profile picture, represented by a grey circle with a white person icon.

Figura 3. Output del fitxer `profile.html`

Seguidament, l'usuari pot accedir a l'apartat **"Departments"** (Figura 4), on hi troba una descripció de la feina que es fa en cada departament i un botó que et redirigeix a la pàgina oficial de Sant Joan de Déu per poder saber més informació. A més a més, sota de cada departament, s'indica el username dels usuaris que tenen accés en cada base de dades (en cas de ser administrador el username apareixerà en tots els departaments).

Seguint navegant pel menú entrem en la secció de l'hospital on tenim l'opció de visualitzar tots els pacients emmagatzemats, o l'opció d'afegir un pacient.

Si l'usuari accedeix a l'apartat **"All patients"** (Figura 5),

L'usuari serà redirigit i podrà veure tots els pacients registrats. Podrà buscar a través d'un filtre per nom del pacient (Nom), Número d'Història Clínica (NHC) o Número d'Història Clínica Familiar (NHCfam), entre d'altres. La taula mostrarà els pacients retornats per la queryset amb la informació bàsica per cada pacient, indicant el nom, NHC, NHCfam, PedFam, Etnia, Gender, i les dades fenotípiques DOB (Date of Birth), AntFam (antecedents familiars) i Consang (consanguinitat).

El filtre que s'aplica fa una cerca exhaustiva i retorna tots aquells documents que continguin el filtre introduït en el formulari, és a dir, no cal indicar exactament el valor esperat, especificant 11 mostrarà tots aquells pacients que continguin un 11 en el NHC. Per no carregar totes les dades en un mateix bloc, s'ha creat una paginació per poder visualitzar els pacients en diferents pàgines anant endavant i endarrere, o accedint directament a la primera o la última taula de pacients.

Departments

Genetics



We perform genetic and genomic analyses for the diagnosis and genetic counselling of our patients.

[See more >](#)

Users:

- guerau
- joe
- laura
- joe

Oncology



We are a national reference centre for the care and research of developmental cancer and an international reference centre.

[See more >](#)

Users:

- laura
- joe

Pediatric hematology



We use diagnostic techniques that allow us to offer a faster and more accurate diagnosis, as well as access to new personalized treatments.

[See more >](#)

Users:

- laura
- joe

Figura 4. Output del fitxer departments.html

Genetics Department

Patients filter

Nom

Ana

Nhc

11

Nhcfam

Pedfam

Etnia

Gender

Search

Patients

#	Nom	NHC	NHCfam	PedFam
View	Anaya Escribano, Ona	996112	nan	nan
View	Amor Aller, Jana	1179388	nan	nan
View	Barón Rivera, Ana	1137745	nan	nan
View	Samper Granados, Miguel	1611492	nan	nan
View	Clemente Garmendia, Jana	1139225	nan	nan

Page 1 of 6 [next](#) [last >](#)

Figura 5. Output del fitxer allPatients.html

Quan l'usuari clica el botó "View", aquest és redirigit a la

pàgina per veure informació més específica del pacient seleccionat (Figura 6) i de totes les mostres (Figura 7). A la part superior de la pàgina es pot veure el nom complet del pacient i seguidament s'hi troben dues seccions.

Leiva Montserrat, Ana

Patient information

NHC

1154519

Gender

1

SampleType

Etnia

nan

Pheno Data

DOB

Sept. 9, 2022, midnight

AntFam

.

Consang.

.

Pheno History

SigneGuia	AltresPheno	AltresPhenoNeg	GeneList	Indicacio	ClinicalOnsetSG
-	HP:0000308, etc	-	-	-	-

Figura 6. Output del fitxer viewPatient.html (Patient information)

Samples

NGS technique

#	PoolNum	SampleName	Afectada	IDSamp	Labels	NHCfam	Onset
View	Pool_281	33PEXSSB035	Si	003/1.110.0241	Main	-	DI se

Page 1 of 1

Figura 7. Output del fitxer viewPatient.html (Samples)

La primera secció mostra tota la informació del pacient, incloent les dades fenotípiques que no varien i l'històric de les dades fenotípiques que poden ser actualitzades al llarg del temps.

En la segona secció s'hi troben totes les proves emmagatzemades que se li han realitzat al pacient. Com que un pacient es pot haver realitzat mostres tant de la tècnica NGS com d'array CGH, aquest pàgina mostra una taula amb totes les mostres realitzades per cada tècnica. En la Figura 7 veiem una taula amb totes les mostres realitzades amb la tècnica NGS, indicant el nom de la mostra (SampleName), la data de petició (PetiDate), la prioritat (Prioritat), el servei (Servei), el tipus de tècnica (Omic), si està informada o no (Informed) o la informació genètica (CNV i SNV), entre d'altres.

Per visualitzar la informació genètica de CNV i SNV l'usuari ha de clicar el botó "View" i serà redirigit a la pàgina per mostrar informació genètica de la mostra seleccionada (Figura 8). L'usuari primer visualitzarà la

informació de la mostra seleccionada, és a dir, veurà el PoolNum i el SampleName, i informació tècnica, és a dir, la màquina amb que s'ha analitzat la mostra (MachineName), el tipus de prova (Prova), el nom de la prova (NomProva) i el proveïdor (Proveïdor), entre d'altres. Seguidament hi trobarà dues taules, una amb totes les CNVs i una altra taula amb les SNVs.

Sample information

PoolNum	Pool_281
SampleName	33PEXSSB035
ProcessError	No
Prova	CES
MachineName	NextSeq500
NomProva	CCP17
TechSeq	Illumina
TargetEnrichment	Captured-based
SeqService	HSJD
Proveïdor	Agilent

CNVs

Informed variant	CHROM	START	END	Length	varType	geneList	R
Inform	chr1	110230428	1102335974	5547	DEL	GSTM1	1
Inform	chr16	70161085	70178524	17440	DUP	CLEC18C PDPR	1
Inform	chr4	69403277	69434284	31008	DUP	UGT2B17	1
Inform	chr6	32485438	32498036	12599	DEL	HLA-DRB5 HLA-DRB1	4
Inform	chr6	31994685	31998333	3649	DEL	C4B_2	5

Page 1 of 4 next last »

Figura 8. Output del fitxer viewSample.html (SNV)

En el cas de voler informar una mostra, l'usuari només haurà de fer un clic al botó "Inform" i directament serà redirigit a la pàgina per informar variants (Figura 9). A través del formulari hi podrà afegir les variants reportades, i clicant al botó "Submit" aquestes seran automàticament introduïdes a la base de dades.

Inform variants

Segregacio:	Validacio:
SosPatHer:	TechValidacio:
ClassificacioLab:	ComfVal:

Submit

Figura 9. Output del fitxer informVariant.html

Per altre banda, si l'usuari en comptes de visualitzar els pacients vol afegir-ne de nous, navegant a través del menú, podrà anar a l'apartat "Add patient" (Figura 10) i inserir diferents fitxers per tal d'actualitzar la base de dades amb les noves dades dels pacients.

Genetics Department

Patient

Seleccionar archivo Ninguno archivo selec. Enviar

Sample CNV

Seleccionar archivo Ninguno archivo selec. Enviar

Sample SNV

Seleccionar archivo Ninguno archivo selec. Enviar

Sample Pheno

Seleccionar archivo Ninguno archivo selec. Enviar

Figura 10. Output del fitxer addPatient.html

En cas de tenir permisos d'administrador i voler gestionar els usuaris, a través del menú es pot accedir a l'apartat "Settings" que automàticament redirigirà a l'usuari a la pàgina d'administració de Django (Figura 7 de l'apèndix).

Finalment, per tancar la sessió, l'usuari podrà clicar el botó "logout" situat a la part superior de la pàgina i abandonar la website (Figura 8 de l'apèndix).

4.3 Visualització de les dades

Com s'ha pogut veure en la pàgina del dashboard (Figura 2), s'han implementat diferents gràfiques per poder visualitzar dades estadístiques les mostres realitzades.

Per tal de tenir una visualització sincronitzada amb la base de dades, s'han implementat els gràfics amb JavaScript per així poder visualitzar les dades en temps real i actualitzades al moment.

Per una banda, s'han creat sis gràfics que donen una visió global d'estadístiques de les diferents mostres realitzades per la tècnica NGS. Al no tenir les dades de la tècnica a-CGH, no s'ha pogut implementar tots els gràfics desitjats fent una comparativa de tècniques.

El gràfic circular de la Figura 11 permet visualitzar el número de mostres de la tècnica NGS realitzades per cada prova. Amb els colors podem veure el nom de la prova (NomProva), i la mida dels triangles mostra el volum de mostres realitzades. Es pot veure com la prova TSOex ha estat la més realitzada, ja que és la prova amb la qual s'han fet més mostres per pacients. Gràcies a aquesta visualització, el departament de genètica podrà saber la diversitat i el volum de proves que han analitzat.

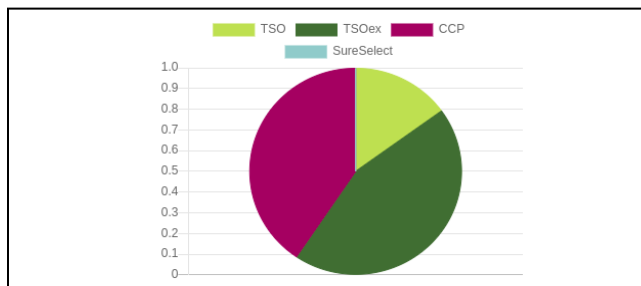


Figura 11. Número de mostres (samples) de la tècnica NGS per cada prova (NomProva)

El gràfic de barres de la *Figura 12* permet veure el número de mostres de la tècnica NGS realitzades per cada servei. Gràcies a la gràfica es pot extreure que els serveis de "Genètica Clínica" i "Neurologia" són els serveis pels quals s'han realitzat més mostres a pacients. El motiu pel qual es mostra la gràfica és saber els serveis que demanen proves al departament de genètica. Saber la quantitat de mostres per servei, dona una idea del tipus de malalties que s'estan intentant diagnosticar.

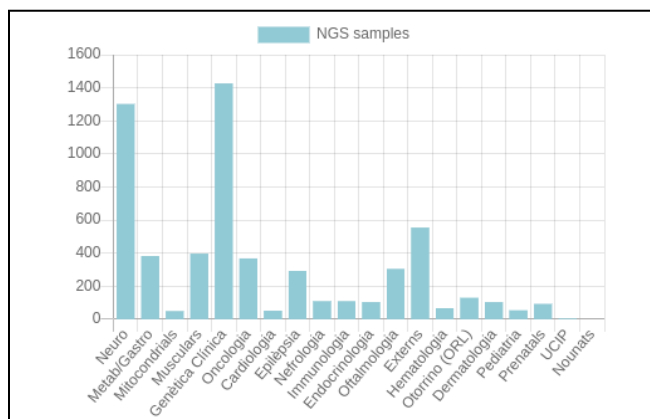


Figura 12. Número de mostres (samples) de la tècnica NGS per servei

La *Figura 13* mostra un gràfic circular amb el número de mostres per tipus de teixit (SampleType). Com que la gran majoria de mostres s'han realitzat per sang perifèrica, s'ha decidit no mostrar-les per així poder visualitzar i comparar els altres tipus de teixit. Es veu clarament que el segon tipus de mostra més realitzada ha estat per teixit fetal, és a dir, fetal líquid amniòtic. A partir d'aquesta visualització es pot saber el tipus de mostres que s'analitzen al departament. Per cada tipus de mostra s'ha d'optimitzar el protocol d'extracció del DNA, per tant la gràfica dona idea de la diversitat de protocols que s'han posat a punt i dels tipus de teixits pels quals tenen la capacitat de fer anàlisis genòmic.

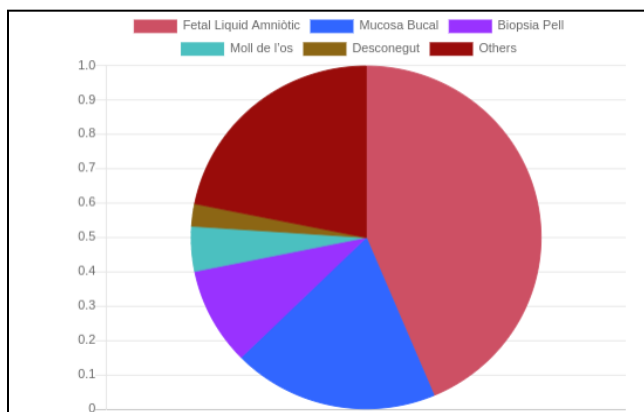


Figura 13. Número de pacients per SampleType

El gràfic logarítmic de la *Figura 14*, mostra la temporalitat de les mostres NGS, és a dir, per cada data de seqüenciació, mostra els diferents anys, i el número de mostres realitzades en aquest. D'aquesta manera es pot veure en el temps si hi ha hagut un augment de proves analitzades pel departament de genètica i saber així la

tendència futura.

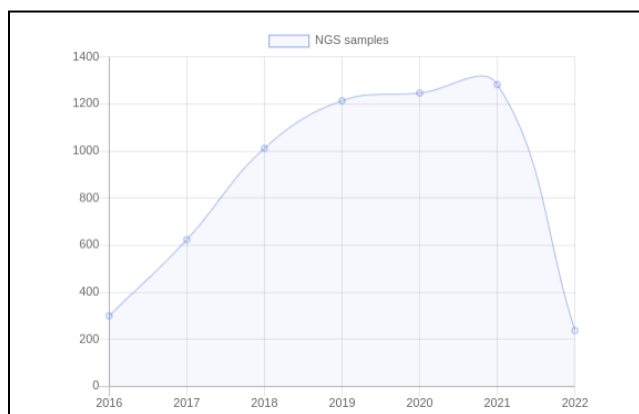


Figura 14. Número de mostres (samples) de la tècnica NGS per any de seqüenciació (PoolDate)

La *Figura 15* mostra un gràfic de barres on s'hi pot veure en l'eix de les y el número de mostres de la tècnica NGS, i a l'eix de les x el tipus de familiar associat al pacient (PedFam). Per fer anàlisis familiars és important tenir controls i saber quin grau de parentiu tenen.

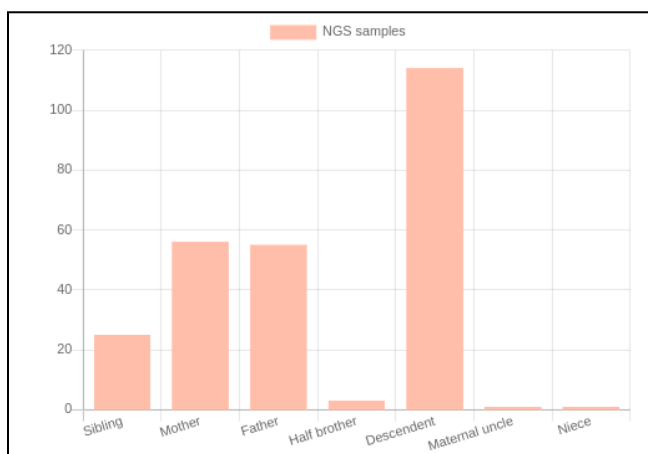


Figura 15. Número de mostres (samples) de la tècnica NGS per relacions familiars (PedFam)

Per últim, s'ha mostrat un gràfic circular per tal de visualitzar la quantitat de mostres afectades, i no afectades. Però al tenir poques dades, la majoria de mostres han estat afectades i n'hi ha molt poques que no ho estiguin (*Figura 9 de l'apèndix*).

4.4 Anàlisis de dades

Un cop creada la plataforma per volcar, estructurar i visualitzar les dades genètiques, he realitzat un anàlisis de les dades per tal de respondre a la hipòtesis que: "Tenint en compte la data de petició de la mostra, la data de seqüenciació i la prioritat, el temps de resposta desde que es sol·licita fer la mostra fins que es seqüencia és de pocs dies en cas de ser prioritària, i de més dies en cas de no tenir tanta prioritat".

Els passos a seguir per desenvolupar l'anàlisis són:

1. Definir els objectius desitjats:
 - Trobar la relació entre la data de petició i data de seqüenciació d'una mostra tenint en compte la

- prioritat (nan, 0, 0-1, 2 i 3) d'aquesta.
 - Mostrar un llistat dels mesos que hi ha hagut un increment de prioritats.
 - Predir l'increment de prioritats futura.
2. Seleccionar les mètriques:
 - Datasets: base de dades del departament de Genètica de l'Hospital Sant Joan de Déu.
 - Mètriques: ProcessError, PetiDate, SeqDate, Prioritat i el càlcul de la mitjana del temps de resposta (la resta entre PoolDate i PetiDate). Les mostres no seqüenciades, és a dir, aquelles mostres marcades amb "ProcessError=No", no s'hauran de tenir en compte per fer la mitjana, ja que significa que s'ha sol·licitat però no s'ha seqüenciat.
 3. Resultats i anàlisi de dades:

La Figura 16 mostra la mitjana del temps que tarda una mostra desde que es sol·licita fins que es seqüencia depenent de la prioritat. Per altre banda, la Figura 17 permet visualitzar en clusters de colors les diferents prioritats, i amb el tamany dels cercles es pot veure que el temps de resposta és menor com major és la prioritat (sent 0 màxima prioritat), és a dir, els cercles són més petits i per tant significa que han passat pocs dies entre la data de petició i la data de seqüenciació.

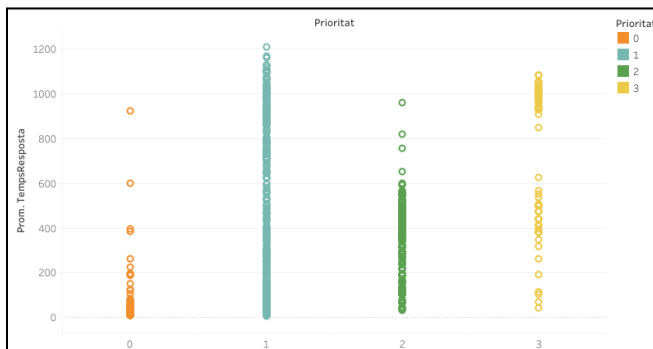


Figura 16. Temps de resposta segons la prioritat de la mostra

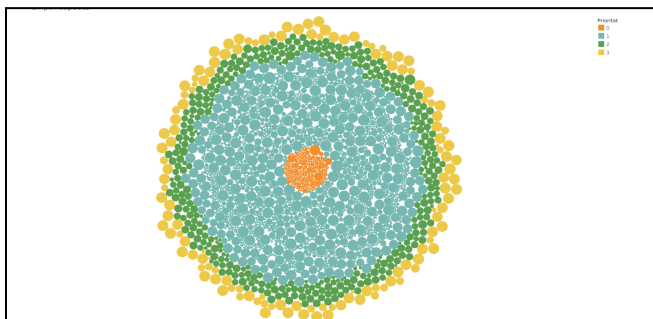


Figura 17. Temps de resposta (tamany) segons la prioritat de la mostra (color)

La Figura 18 mostra el temps de resposta per cada prioritat en cada mes de seqüenciació de la mostra. D'aquesta manera es pot veure en el temps si hi ha hagut un augment del número de mostres analitzades per cada prioritat. En cada cas hi ha un creixement o decreixement diferent, però la major variació en el temps és la prioritat 1 (línia marcada en blau). A partir del mes de seqüenciació

de la mostra i la prioritat d'aquesta, podem dir que l'últim mes hi ha hagut un decreixement de les prioritats, segurament per falta de dades emmagatzemades.

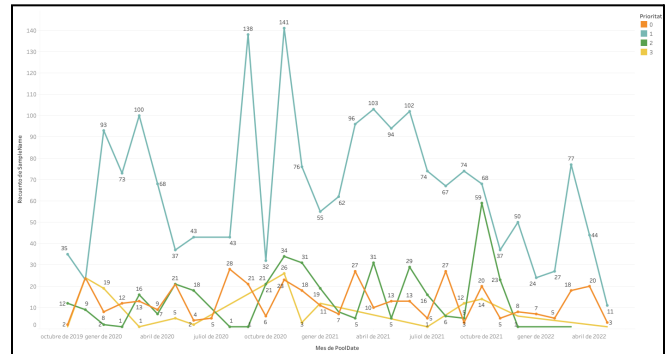


Figura 18. Temps de resposta per prioritat i mes de seqüenciació (PoolDate)

Per últim, per tal de predir si hi haurà un increment o no de les prioritats el següent mes, s'ha creat una gràfica que mostra una línia de tendència per cada prioritat (Figura 10 de l'apèndix). La línia de tendència permet veure com la majoria de prioritats tendiran a mantenir-se constants.

5 TESTS

Amb l'objectiu de comprovar que la base de dades ha estat creada correctament i no té incoherències, s'han plantejat diferents consultes per testejar les relacions entre les dades.

Algunes d'aquestes consultes són:

- Trobar el pacient que es correspon a la mostra SampleName = "01PEX1C1" i PoolNum = "Pool_15NS_DN". Primer s'ha hagut de cercar la mostra indicada i seguidament seleccionar la variable "patient" per extreure el ID del document del pacient i poder retornar les seves dades. D'aquesta manera s'ha comprovat que la connexió entre les col·leccions "Samples" i "Patients" sigui correcta.
- Retornar totes les mostres realitzades amb la tècnica NGS del pacient amb _id = "62b3358adc8ebf59d28b7cad". Primer s'ha hagut de cercar el pacient per l'ID de l'objecte, i després cercar a la col·lecció "Samples" els IDs dels objectes (samples) emmagatzemats en un array a la variable SamplesNGS del pacient seleccionat.
- Trobar quantes mostres s'han realitzat a través del fetus. S'ha hagut de filtrar i retornar tots els documents on la variable SampleType comencés per la lletra "F_", la qual indica que la mostra s'ha fet a través del fetus. Seguidament s'ha fet un recompte del número de mostres retornat.

Per altra banda, per verificar que l'aplicació compleix amb els requisits indicats pels responsables del departament de genètica de l'Hospital Sant Joan de Déu, s'han realitzat reunions setmanals per testejar i fer una demo de la website, i poder així en cas d'haver-hi algun error, ser a temps a modificar-lo. Els tres usuaris creats tenen diferents permisos, l'usuari Laura i Joe tenen el rol

d'administradors, en canvi l'usuari Guerau, té permisos restringits. Com que té el rol de staff, només té accés a la base de dades del seu departament, en aquest cas de genètica.

Per testejar aquestes funcionalitats, s'ha accedit i fet proves amb els tres perfils. A més a més per provar que el filtratge dels pacients sigui correcte, s'han buscat diferents pacients amb característiques diverses, i s'ha comprovat amb els fitxers rebuts pel departament que les mostres corresponen realment amb el pacient seleccionat.

6 CONCLUSIÓ I MILLORES

- S'ha dissenyat i implementat una website per a poder volcar, estructurar, visualitzar i analitzar les dades genètiques dels pacients.
- S'ha dissenyat una base de dades pel departament de Genètica.

Com a millores:

- Que les ontologies definides juntament amb el projecte de la Marató de TV3 siguin visualitzades a la plataforma amb la transformació per ser enteses pels doctors del departament de genètica.
- Crear una base de dades per cada servei.
- Posar en funcionament la pàgina d'afegir pacients i d'informar variants

AGRAÏMENTS

Al meu tutor i en especial al Dr. Guerau Fernandez i al Joe Kaen per la seva ajuda, paciència i dedicació.

BIBLIOGRAFIA

- [1] Guía Metabólica, Hospital Sant Joan de Déu, Los registros de pacientes son una herramienta importante para la investigación, <https://metabolicas.sjdhospitalbarcelona.org> [Consulta: 10 de març del 2022]
- [2] Balsamiq Wireframes, <https://balsamiq.com> [Consulta: 19 de març del 2022]
- [3] MoonModeler, Data modeling tool, <https://www.datensen.com> [Consulta: 25 de maig del 2022]
- [4] MongoDB, <https://www.mongodb.com> [Consulta: 20 de maig del 2022]
- [5] Django REST Frame., <https://www.django-rest-framework.org> [Consulta: 22 de maig del 2022]
- [6] MongoDB, Connect Django and MongoDB Using Djongo, <https://www.mongodb.com> [Consulta: 19 d'abril del 2022]
- [7] MongoDB, Connect Django and MongoDB Using MongoEngine, <https://www.mongodb.com> [Consulta: 19 d'abril del 2022]
- [8] MongoDB Compass, <https://www.mongodb.com> [Consulta: 2 de juny del 2022]
- [9] Tableau, Business Intelligence and Analytics Software, <https://www.tableau.com> [Consulta: 15 de juny del 2022]
- [10] ShinyProxy, <https://www.shinyproxy.io/> [Consulta: 10 de març del 2022]
- [11] GitHub Docs, <https://docs.github.com> [Consulta: 15 d'abril del 2022]
- [12] NIH, National Cancer Institution, Definition of CNV, <https://www.cancer.gov> [Consulta: 19 de març del 2022]
- [13] NIH, National Cancer Institution, Definition of SNV,

<https://www.cancer.gov/publications/dictionaries/genetics-dictionary/def/snv> [Consulta: 19 de març del 2022]

- [14] Beacon Network, <https://beacon-network.org> [Consulta: 26 de març del 2022]
- [15] Pandas, <https://pandas.pydata.org> [Consulta: 18 de maig del 2022]
- [16] Pymongo, <https://pypi.org> [Consulta: 1 d'abril del 2022]
- [17] Scitable by Nature Education, A collaborative Learning Space for Science, Microarray-based Comparative Genomic Hybridization (aCGH), <https://www.nature.com> [Consulta: 24 de març del 2022]
- [18] Illumina, Introduction to NGS, <https://www.illumina.com> [Consulta: 24 de març del 2022]
- [19] GitHub, Insert-Data repository, Collate and transform data and export patients and samples to MongoDB Server, <https://github.com/LaPoEs/Insert-Data> [Consulta: 27 de juny del 2022]
- [20] GitHub, HSJD repository, Django app for genetics data visualization for Hospital Sant Joan de Déu, <https://github.com/LaPoEs/HSJD> [Consulta: 27 de juny del 2022]

APÈNDIX

A1. DISSENY - DISSENY DE LA INTERFÍCIE

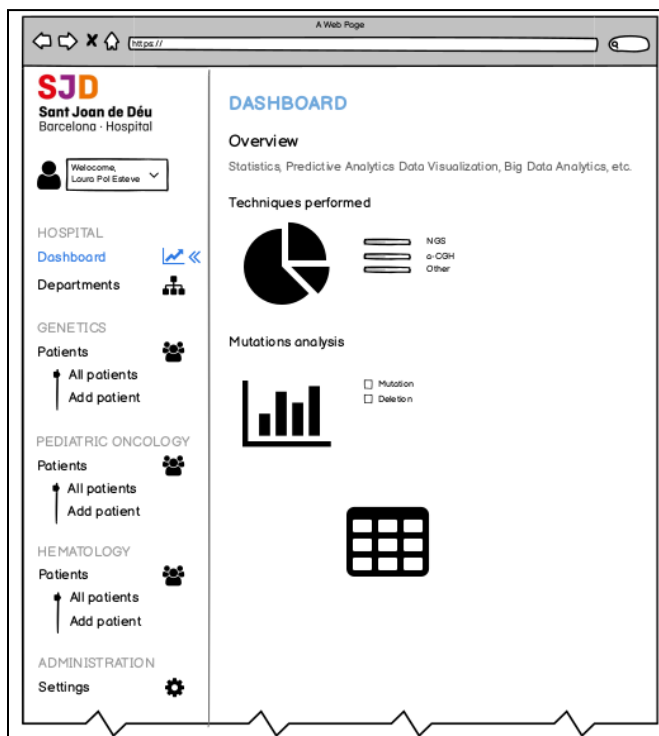


Figura 1. Wireframe de la pàgina "Dashboard"

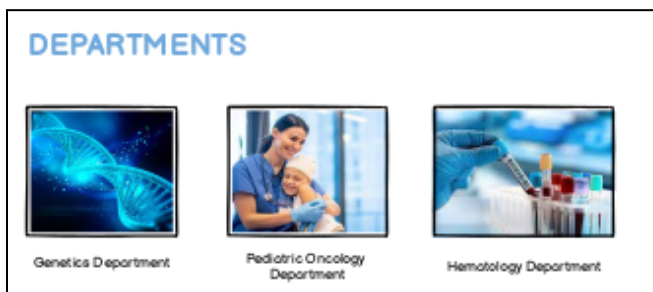


Figura 2. Wireframe de la pàgina "Departments"

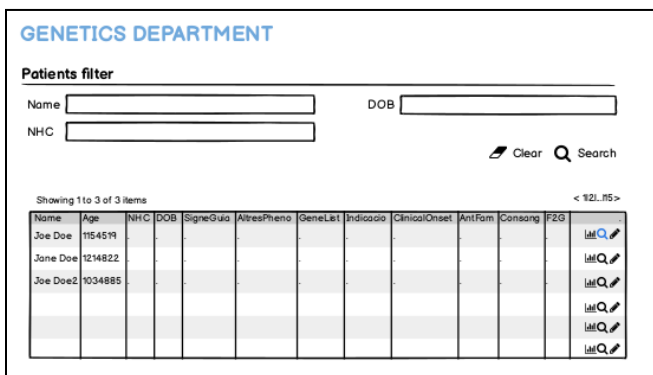


Figura 3. Wireframe de la pàgina "All patients"

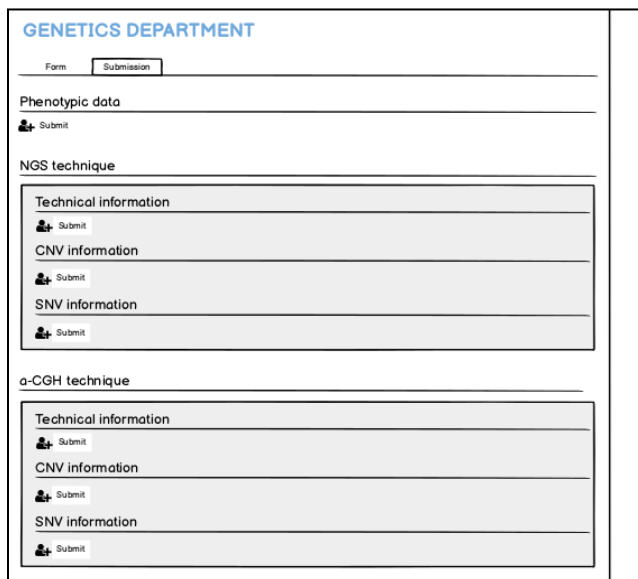


Figura 4. Wireframe de la pàgina "Add patient"

A2. DESENVOLUPAMENT - IMPLEMENTACIÓ DE LA WEBSITE

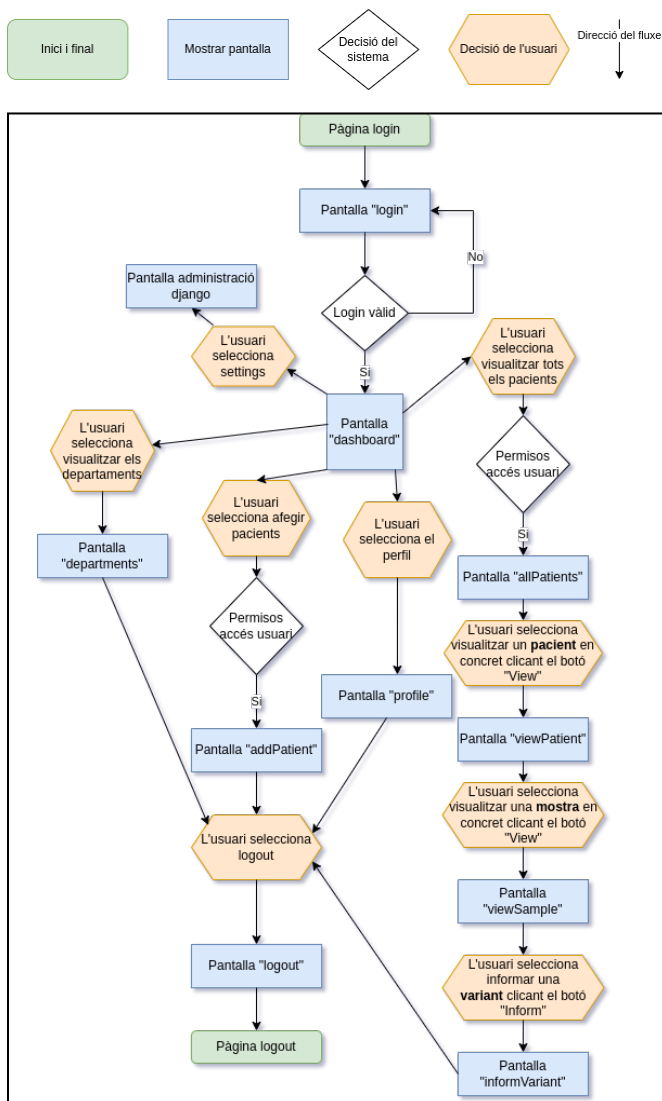


Figura 5. Diagrama de flux

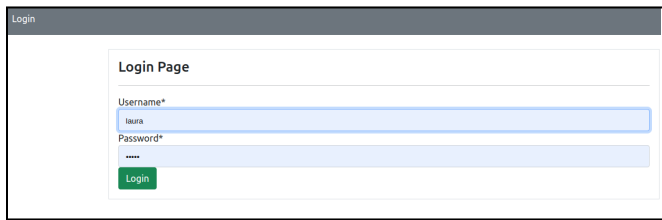


Figura 6. Output del fitxer login.html

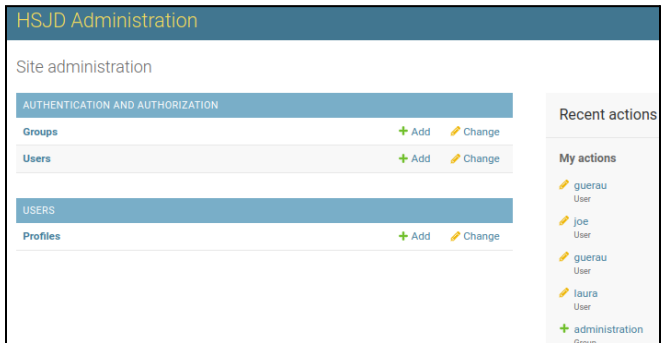


Figura 7. Pàgina d'administració de django

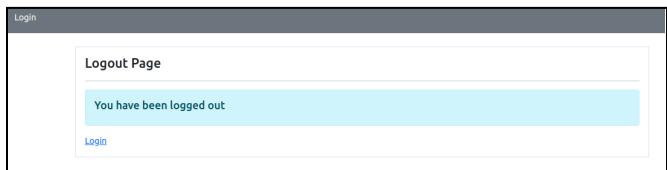


Figura 8. Output del fitxer logout.html

A3. DESENVOLUPAMENT - VISUALITZACIÓ DE LES DADES

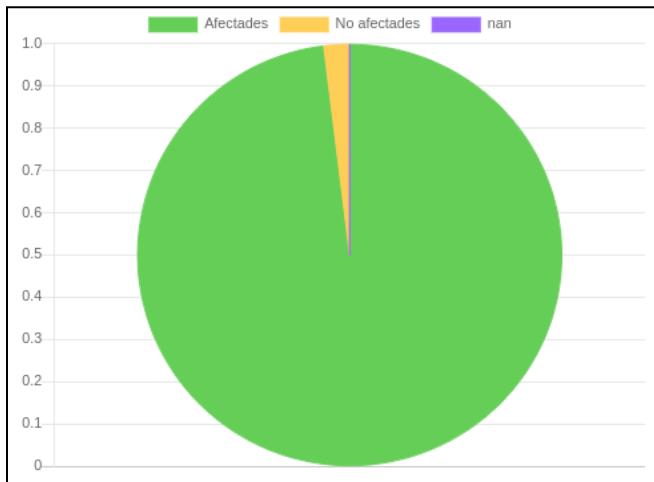


Figura 9. Número de mostres (samples) per afectació

A4. DESENVOLUPAMENT - ANÀLISI DE LES DADES

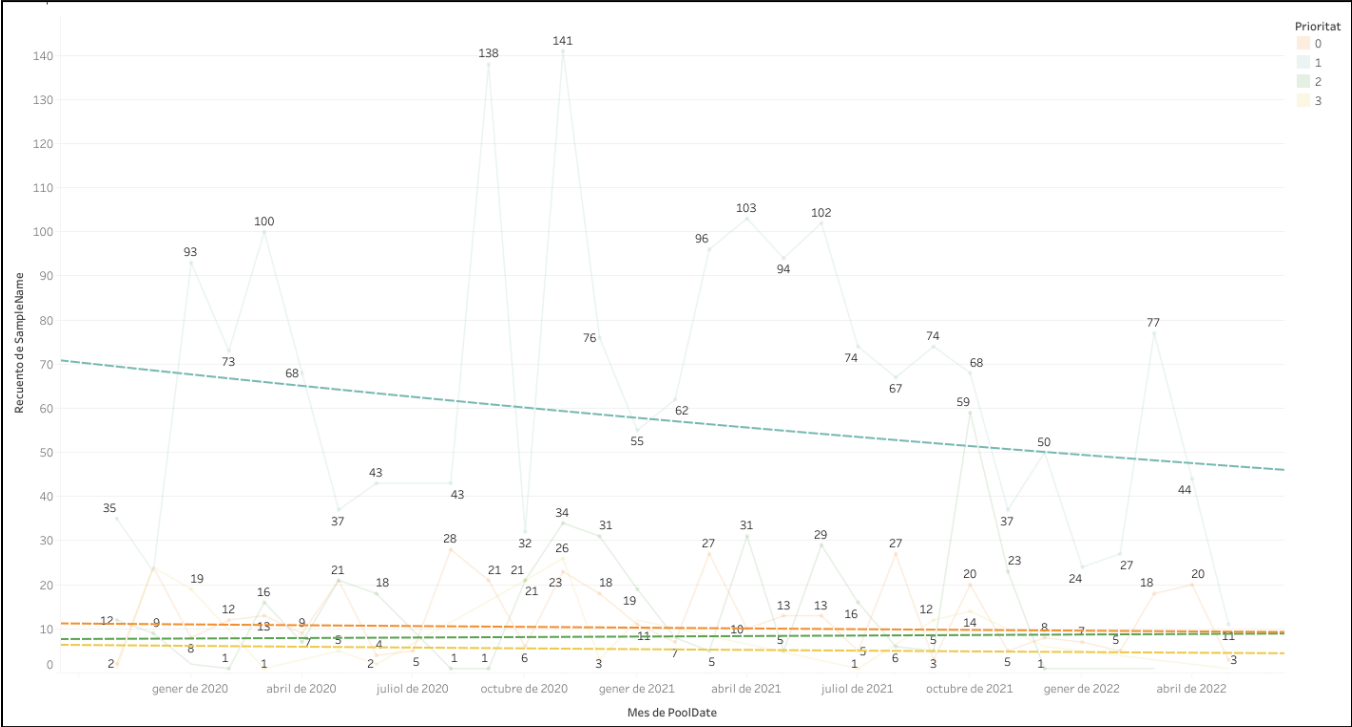


Figura 10. Tendència de la quantitat de mostres per mes de seqüenciació i prioritat