

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento della Salute della Donna e del Bambino

SCUOLA DI DOTTORATO DI RICERCA IN: Medicina dello Sviluppo e Scienze della
Programmazione Sanitaria

INDIRIZZO: Scienze della Programmazione Sanitaria

CICLO XXVIII

**La Cartella Clinica Informatizzata nel percorso diagnostico-assistenziale del Malato
Raro: sviluppo e implementazione di un sistema di raccolta e analisi
dell'informazione clinica fenotipica e genotipica**

*Computerized Medical Record in the diagnostic-care of the Rare Diseases: development
and implementation of a system for the collection and analysis of clinical phenotypic
and genotypic data*

Direttore della Scuola: Ch.mo Prof. Giuseppe Basso

Coordinatore d'indirizzo: Ch.mo Prof. Paola Facchin

Supervisore: Ch.mo Prof. Prof. Paola Facchin

Dottorando : Dott.ssa Ilaria Festa

2015-2016

A Paolo

INDICE

RIASSUNTO	7
ABSTRACT	13
1 INTRODUZIONE	19
1.1 LE MALATTIE RARE: DEFINIZIONE	19
1.2 PROFILO EPIDEMIOLOGICO DELLE MALATTIE RARE	20
1.3 SISTEMI CLASSIFICATIVI PER LE MALATTIE RARE	25
1.3.1 DESCRIZIONE DEL FENOTIPO	26
1.3.2 DESCRIZIONE DEL GENOTIPO	34
1.4 SISTEMI DI MONITORAGGIO SANITARIO	47
1.4.1 REGISTRI PER LE MALATTIE RARE	47
1.4.2 FASCICOLO SANITARIO ELETTRONICO INDIVIDUALE	59
1.5 I MALATI RARI SENZA DIAGNOSI	65
1.5.1 SCENARIO DI RIFERIMENTO	65
1.5.2 UNDIAGNOSED DISEASES PROGRAM IN USA	67
1.5.3 GLI SCENARI DI RICERCA	70
1.5.4 I PROBLEMI COMUNI	76
2 OBIETTIVI	79
2.1 CREAZIONE DI STRUMENTI DI DESCRIZIONE ANALITICA E PRECODIFICATA DELL'INSIEME DI DIVERSI FENOTIPI E GENOTIPI	79
2.2 TRASFERIMENTO DEGLI STRUMENTI DESCRITTIVI IN PRODOTTI UTILI PER LA GESTIONE CLINICA	80
2.3 IMPLEMENTAZIONE DEI PRODOTTI IN DIVERSI CONTESTI REALI DI ASSISTENZA CLINICA D'ECCELLENZA	81
3 MATERIALI E METODI	83
3.1 IL DISEGNO SPERIMENTALE	83
3.2 LE VARIABILI E LA LOGICA CLASSIFICATORIA	84
3.3 ENTITÀ E RELAZIONI	85
3.4 LE FONTI	86
3.5 I CONTENUTI ED I PRODOTTI	88

3.6	L'IMPLEMENTAZIONE	89
3.7	I METODI DI ANALISI	90
4	RISULTATI	93
4.1	CONTENUTI DEL MODULO CCI	93
4.2	CONTENUTI DEL DATABASE RELAZIONALE	97
4.2.1	ENTITÀ E RELAZIONI	97
4.2.2	PAZIENTI	100
4.3	IL SISTEMA ESPERTO	108
5	DISCUSSIONE E CONCLUSIONI	111
6	BIBLIOGRAFIA	115
	APPENDICI	123
	INDICE DELLE FIGURE	123
	INDICE DELLE TABELLE	124

RIASSUNTO

A livello internazionale non esiste una definizione univoca di malattia rara: paesi europei ed extraeuropei la stabiliscono in base a differenti soglie di prevalenza in accordo alle legislazioni vigenti: nell'Unione Europea (UE), una patologia è rara se presenta una prevalenza di popolazione uguale o inferiore a 1/2000 abitanti (0,05%). Le malattie rare sono, sotto il profilo clinico-epidemiologico, un gruppo molto eterogeneo, con circa 5000-8000 entità per lo più ad origine genetica, spesso gravi, cronicamente invalidanti, progressive e potenzialmente fatali, che rappresentano un grave problema di salute pubblica; nonostante ciò, sono state in passato di scarso interesse scientifico, sono poco incluse nelle classificazioni internazionali e poco visibili a livello d'informazione sanitaria, per molte di esse inoltre non sono ancora disponibili dati epidemiologici. La ragione principale sembra essere l'assenza di un sistema di corretta codifica e rintracciabilità nei sistemi informativi sanitari: gli attuali sistemi classificativi e di raccolta dell'informazione fenotipica e genotipica sono ampiamente eterogenei, privi di terminologia standard, scarsamente inter-operabili, unicamente dedicati alla classificazione del fenotipo, o con caratteristiche di database genomici con profili fenotipi variante-correlati. Alcuni di questi database, sono già operativi, altri in via di sviluppo, di pari passo con l'implementazione delle tecnologie NGS, che hanno garantito alla comunità medico scientifica, rispetto alle tecniche tradizionali, lo studio di una vastissima gamma di malattie genetiche, con crollo dei costi e dei tempi d'analisi dei dati. A livello mondiale sono nati centri e progetti di ricerca altamente specializzati in tecniche di sequenziamento genomico e nello sviluppo di database di varianti genomiche, che hanno reso disponibile una grande varietà e quantità di nuove informazioni, spesso di difficile gestione ed interpretazione. L'imperativo della comunità scientifica è pertanto l'organizzazione sistematica di questa enorme quantità di dati, potenzialmente utili alla corretta interpretazione del significato patogeno delle varianti genomiche in relazione al fenotipo clinico. In base a dati emergenti dalla letteratura, i registri di pazienti e di patologia costituiscono strumenti interessanti per lo sviluppo di conoscenze utili sotto il profilo epidemiologico-clinico in ambito di malattie rare. Molti registri esistenti devono affrontare problematiche di qualità, standardizzazione e accessibilità dei dati, di ottimizzazione delle risorse e di

interoperabilità legate alla frammentazione dei sistemi di codifica delle malattie. La situazione a livello europeo ed extra-europeo sembra favorire la costituzione di registri globali o di piattaforme comuni di coordinamento di tutti i registri esistenti e futuri, seguendo un approccio uniforme, coordinato e di alta qualità. La realizzazione di questo modello è attualmente un obiettivo strategico della comunità europea nell'ambito del progetto EPIRARE e della comunità statunitense nell'ambito del progetto sul Global Rare Diseases Registry (GRDR). Nel percorso evolutivo che sostiene l'implementazione del servizio sanitario pubblico per migliorare l'esercizio del diritto alla salute del cittadino è fondamentale l'attuale sviluppo di tecnologie informatiche innovative che rendano disponibile informazioni complete, condivise ed idonee a garantire al cittadino la migliore continuità assistenziale. In Italia Fascicolo Sanitario Elettronico ("FSE") è la raccolta informatizzata dei dati della storia clinica di un individuo, progettata nell'ambito del Servizio Sanitario Nazionale (SSN) secondo le Linee Guida Nazionali emanate dal Ministero della Salute nel novembre 2010. Il processo è lento, ma con l'introduzione del FSE è auspicabile che nel prossimo futuro, si assista ad un rivoluzionario cambiamento dei rapporti tra infrastruttura sanitaria e cittadino, con miglioramento del flusso di dati e del processo diagnostico-assistenziale. Anche nel contesto internazionale è documentata l'utilità del cosiddetto "Electronic Health Record" (EHR) o "Personal Health Record" (PHR) a supporto di tale processo in ambito sanitario, oltre che per la misurazione ed il monitoraggio della qualità delle cure e per la ricerca scientifica traslazionale, ma realizzazione e lo sviluppo del FSE sono ancora in fase preliminare e le sfide da affrontare per rendere applicativa tale infrastruttura sono numerose, soprattutto per i paesi in via di sviluppo.

In ambito di malattie rare altra sfida importante su molti fronti per i sistemi sanitari è quella rappresentata dal gruppo eterogeneo dei pazienti senza diagnosi, ovvero gli individui con malattie note non correttamente riconosciute, che subiscono ritardi ed errori diagnostici, o gli individui con patologie sconosciute; questi pazienti sovrautilizzano percorsi diagnostico-terapeutici inadeguati e presentano bisogni assistenziali non soddisfatti, con elevati costi umani e sociali. Ad oggi il problema dei pazienti senza diagnosi è stato universalmente riconosciuto a livello mondiale come centrale nella programmazione sanitaria della rete di assistenza per malati rari, ma appare ancora irrisolto: i tentativi messi in atto in USA con Undiagnosed Diseases Program (UDP) del

National Institute of Health (NIH) Clinical Center hanno dato risultati promettenti, ma sono apparsi ancora scarsamente sostenibili nella pratica clinica e di ancora limitato valore epidemiologico.

A livello nazionale alcuni Centri Accreditati per eccellenza nella diagnosi e cura di specifiche malattie rare o gruppi di malattie rare correlate, hanno istituito progetti mirati alla ricerca di soluzioni metodologiche per raccolta dell'informazione clinica nella sua corrispondenza fenotipo-genotipica, a sostegno del processo diagnostico-assistenziale dei malati rari, con particolare riguardo ai pazienti senza diagnosi o con un generico sospetto diagnostico: Progetto di ricerca Regione-Università "*Next Generation Sequencing and Gene Therapy to Diagnose and Cure Rare Diseases in Regione Emilia Romagna (RARER)*" e Progetto Ministeriale "*A multicenter collaborative research network for the identification and study of rare undiagnosed patients: the impact on the rare disease National Health Service network (UnRareNet)*".

Nel più ampio contesto dei suddetti progetti di ricerca il presente progetto di dottorato si è posto l'obiettivo primario di realizzare uno strumento unico di raccolta di dati analitici descrittivi il fenotipo e il genotipo di pazienti con diagnosi nota o sospetta, all'interno di gruppi di malattie rare selezionate, basato su una piattaforma informatizzata, con implementazione del modello del Registro Regionale per le Malattie Rare già in uso in Regione Veneto.

Ulteriore obiettivo è stata la creazione di un prodotto specifico utile alla gestione clinica in ambito di malattie rare, soprattutto rispetto ad efficienza, rapidità e correttezza d'inquadramento diagnostico, attraverso il trasferimento dello strumento descrittivo di raccolta dell'informazione clinica fenotipo-genotipica di pazienti con malattie o gruppi di malattie rare selezionate, nel modulo generalizzato di Cartella Clinica Informatizzata (CCI). Il modulo CCI deve possedere caratteristiche trasversali rispetto ai registri di malattia ed alle cartelle informatizzate ospedaliere, dedicate ad un singolo episodio di cura e strutturarsi sul modello emergente dei record sanitari individuali informatizzati, quali ad esempio in Italia, il Fascicolo Sanitario Elettronico individuale.

Il valore aggiunto è l'integrazione delle potenzialità di ricerca, offerte dalla struttura del database relazionale, con le applicazioni cliniche del modulo CCI nella

gestione pratica del percorso assistenziale e di cura dei malati rari. Si è programmato inoltre l'inserimento in rete del modulo CCI, nell'ambito di un utilizzo condiviso dell'informazione clinica, genetica ed assistenziale in contesti reali d'assistenza clinica d'eccellenza, quali quelli offerti dai progetti RARER e UnRareNet.

Il progetto si è strutturato in 3 diverse fasi:

1. identificazione e condivisione, all'interno del protocollo comune RARER, della metodologia idonea alla creazione di 8 registri di malattie rare, con implementazione delle banche dati clinica e strutturazione di un modulo di raccolta del dato genotipico;
2. realizzazione e condivisione in rete del modulo CCI, come prodotto sostenibile in ambito clinico, contenente l'informazione clinica fenotipica e genotipica descritta all'interno di un database relazionale;
3. implementazione sistema d'acquisizione dati e del modulo CCI con l'ampliamento della popolazione nel contesto del progetto UnRareNet, analisi dei dati, con identificazione della correlazione tra quadri fenotipici e genotipo e progettazione di un sistema esperto in grado di gestire procedure complesse di elaborazione dati.

Il progetto di ricerca ha previsto la realizzazione del database relazionale con raccolta di dati clinici di malati rari con diagnosi nota o sospetta nell'ambito di alcune patologie rare di specifico interesse attraverso il modulo CCI. Le entità nosologiche e le entità principali e secondarie del database, descrittive il fenotipo ed il genotipico del paziente sono classificate e quindi organizzate secondo una logica classificatoria nodulare e multigerarchica ad albero, che integra i sistemi classificativi uni e multidimensionali riconosciuti a livello internazionale e contemporaneamente rispetta la suddivisione anatomica e funzionale dell'organismo umano.

Il database contiene 72360 record, distribuiti tra entità principali e secondarie (segni, sintomi, comorbidità, menomazioni, indagini diagnostiche, attributi) e relazioni tra entità, 117 differenti entità nosologiche, 63 differenti geni e 1186 pazienti totali. I risultati ottenuti nella fase di creazione ed implementazione del sistema di raccolta dell'informazione clinica, con la messa in atto del modulo CCI nel contesto del progetto RARER, hanno dimostrato l'usabilità del prodotto in un ambito preliminare;

l'ampliamento dell'applicazione clinica nel contesto del progetto UnRareNet, ha prodotto risultati promettenti ed il sistema ha mostrato una buona tenuta in condizioni di fenotipi simili e dissimili, gestendo in modo completo informazioni cliniche riguardanti complessivamente numerose differenti malattie rare e dimostrandosi quindi generalizzabile, flessibile ed inter-operabile. I dati contenuti sono al momento ancora limitati, ma possedendo caratteristiche analitiche e precodificate appaiono subito utilizzabili per future elaborazioni complesse. La raccolta dell'informazione genetica e l'integrazione con l'informazione clinica sono apparse anch'esse fattibili, con risultati promettenti, seppur nell'ambito del progetto UnRareNet ancora in fase assolutamente preliminare; i pazienti senza diagnosi sono circa il 10% circa del totale, ma di questi più della metà non dispone ancora d'indagine genetica, che si è rivelata utile alla conferma diagnostica nel 53% dei casi con malattia nota. Sul totale dei pazienti del campione l'utilizzo di tecnologie NGS appare in questa fase limitato, ma i risultati ottenuti sottolineano le potenzialità delle moderne tecniche di sequenziamento genomico nell'implementazione del processo diagnostico; l'ampliamento della popolazione di pazienti arruolati e l'applicazione mirata di queste tecnologie potrebbe consentire nuova più ampia conoscenza sul profilo fenotipico e genotipico delle malattie considerate. Disponendo di una più ampia popolazione di pazienti sarà possibile l'istruzione del sistema esperto, di cui si sono gettate le basi nel contesto del presente progetto, attraverso la raccolta di frequenze teoriche di segni clinici e reperti diagnostici di alcuni gruppi di malattie rare correlate (metaboliche, mitocondriali), derivanti dalla letteratura internazionale e destinate al sistema per elaborazione di casi clinici simulati. Il sistema esperto potrà consentire la validazione di diagnosi simulate attraverso la raccolta di nuovi casi con diagnosi nota e la formulazione di nuove diagnosi di patologie note o l'identificazione di nuove entità nosologiche, anche con utilizzo di reti neurali o reti di kohonen e metodologie Fuzzy. La metodologia di raccolta dell'informazione fenotipica e genotipica del paziente con malattia rara, proposta nel presente progetto di dottorato si è dimostrata idonea a soddisfare i requisiti d'applicabilità richiesti ed ha suggerito una serie di applicazioni future che mirano all'ampliamento delle conoscenze sulla variabilità genomica della malattie rare ed alla riduzione del numero dei mali senza diagnosi.

ABSTRACT

Internationally there is no single definition of a rare disease: in European countries and beyond is determined on the basis of different thresholds of prevalence, in accordance with regulatory requirements: in the European Union (EU), a disease is rare if it has a population prevalence of less than or equal to 1/2000 inhabitants (0.05 %). Rare diseases, from a clinical and epidemiological point of view-, are a very heterogeneous group, with about 5000-8000 entities, mostly with genetic origin, often severe, chronically debilitating, progressive and potentially fatal, representing a major public health problem; nevertheless, in the past, have received little scientific interest, are hardly included in international classifications and barely visible at the level of health information, for many of them also epidemiologic data are not yet available. The main reason seems to be the absence of a system of correct coding and traceability in health information systems: the existing classification systems for collecting phenotypic and genotypic data, are widely heterogeneous, lacking a standard terminology, slightly inter-operable, solely dedicated to the classification of the phenotype or supplied with database features of variants and related phenotypes profiles. Some of these databases are already operating, other developing, along with the implementation of NGS technologies, which guaranteed the medical scientific community, compared to traditional techniques, the study of a wide range of genetic diseases, with the collapse of cost and time of data analysis. Worldwide were established centers and research projects in highly specialized techniques of genome sequencing and development of databases of genomic variants, that have made available a large variety and quantity of new information, often difficult to manage and interpret. The commitment of the scientific community is therefore the systematic organization of that huge amount of data, potentially useful to the correct interpretation of the pathogenic significance of genomic variants in relation to the clinical phenotype. Based on emerging data from the literature, the records of patients and pathologies constitute interesting tools for the development of knowledge useful for the clinical and epidemiological aspect and in the field of rare diseases. Many existing registers must address issues of quality, accessibility and standardization of data, optimization of resources, and interoperability, related to fragmentation of coding systems of diseases. The situation on European and

international scene seems to favor the establishment of global registers or common platforms for coordination of all existing and future registers, following a uniform approach, coordinated and of the highest quality. The implementation of this model is currently a strategic objective of the European community within the project EPIRARE and of the US community in the project on Global Rare Diseases Registry (GRDR). In the evolutionary path that supports the implementation of the public health service to improve the exercise of the right to healthcare of citizens, the current development of innovative information technologies, which make available comprehensive information, shared and able to guarantee citizens the best continuity of care, is critical. In Italy, the Electronic Health Record (FSE) is the computerized collection of data of the clinical history of an individual, designed as part of the National Health Service (SSN) according to the National Guidelines issued by the Ministry of Health in November 2010. The process is slow, but with the introduction of the SFE is to be hoped that in the near future, we witness a revolutionary change in the relationship between health infrastructure and citizen, so far characterized by communication barriers, rigid and cumbersome bureaucratic procedures and paperwork with improvement of the flow of data and the diagnostic process and healthcare. Also in the international context is documented the usefulness of the so-called "Electronic Health Record" (EHR) or "Personal Health Record" (PHR) to assist this process in healthcare, as well as for measuring and monitoring the quality of care and for translational research, but the construction and development of the FSE are still at a preliminary stage and the challenges of making the application infrastructure are numerous, especially for developing countries.

Another major challenge on many fronts to health systems in the field of rare diseases is that represented by the heterogeneous group of patients without diagnosis, individuals with known diseases incorrectly recognized, experiencing delays and diagnostic errors, or individuals with unknown diseases; these patients over-use inadequate diagnostic pathways and their care needs are not satisfied, with high human and social costs.

To date the problem of patients without diagnosis was universally recognized worldwide as central, to the planning of health care networks for patients with a rare disease, but its still unresolved: the attempts made in the USA with Undiagnosed

Diseases Program (UDP) of National Institute of Health (NIH) clinical Center have shown promising results, but appeared poorly sustainable in clinical practice and still have limited epidemiological value.

Nationally some Accredited Centres for excellence in the diagnosis and treatment of specific rare diseases or groups of related rare diseases, have established projects targeted on finding solutions to methodological collection of clinical phenotype-genotype correspondence, supporting the diagnostic process of patients with rare diseases, with special attention to patients without a diagnosis or with a generic presumptive diagnosis: Region-University research Project "Next Generation Sequencing and Gene Therapy to diagnose and Cure rare Diseases in Emilia Romagna (rarer)" and Ministerial Project "A multicenter collaborative research network for the identification and study of rare undiagnosed patients: the impact on the rare disease National Health Service network (UnRareNet)".

In the broader context of these research projects present PhD project designed to create a unique collection of analytical data, describing phenotype and genotype features of patients with known or suspected diagnosis, between groups of selected rare diseases, based on a computerized platform, with implementation of the model of the Regional Registry for Rare Diseases, already used in the Veneto Region.

A further objective was the creation of a specific product useful for clinical management in the field of rare diseases, especially with respect to efficiency, speed and accuracy of diagnostic assessment, through the transfer of the instrument for collection of clinical phenotype-genotype data in patients with diseases or groups of selected rare diseases in the form of generalized Computerized Medical Record (CCI).

The form CCI should have features which cut across rare diseases registry and the hospital medical records, dedicated to a single episode of care, and should structure on the emerging pattern of computerized personal health record, such as in Italy, the "Fascicolo Sanitario Elettronico".

The added value is the integration of the relational database research potential with the clinical applications of CCI in the practical management and care of patients with rare diseases. It was also planned the networking of CCI as part of a shared use of

clinical, genetic and healthcare data in real contexts of clinical excellence, such as those offered by the RARER and UnRareNet projects.

The project is structured in three different phases:

1. identification and sharing, within the common RARER protocol, of a suitable methodology for the construction of 8 different disease registries, with implementation of the clinical database and structuring a form of collection of genotypic data;
2. realization and network sharing of module CCI such as a sustainable product in a clinical setting, containing clinical information in a relational database;
3. implementation of the acquisition data system and module CCI, with the expansion of the patient population in the context of the project UnRareNet, data analysis, identifying the correlation between genotype and phenotype pictures, and design of an expert system able to manage complex procedures for data processing.

The research project proposed the implementation of a relational database with the collection through CCI of clinical data in patients with rare diseases with known or suspected diagnosis, in the context of some rare diseases of specific interest.

The clinical entities and major and minor entities, describing phenotype and genotype features are classified in the database according to a classificatory nodular and multihierarchical like a tree method, which integrates uni and multidimensional internationally recognized classification systems and, at the same time respects the anatomical structure and function of the human organism.

The database includes 72,360 records, distributed between the main and secondary entities (signs, symptoms, comorbidity, impairments, diagnostic, attributes) and relationships between entities, 117 different clinical entities, 63 different genes and 1,186 total patients. The results obtained during the development and implementation of the system of collection of clinical information, with the implementation of the module CCI in the context of the project RARER, have proved the usability of the product in a preliminary context; the expansion of the clinical application in the context of the project UnRareNet, has produced promising results and the system showed a good

performance in terms of similar and dissimilar phenotypes, managing fully clinical information regarding total number of different rare diseases and proving thus be generalized flexible and interoperable. The data are still limited at the time, but having analytical characteristics pre-coded, are immediately usable for future complex processing. The collection of genetic information and integration with clinical information appeared feasible with promising results, although the project UnRareNet this is still absolutely at a preliminary stage; patients without diagnosis is about 10% of the total, but more than half of them do not yet have genetic investigation, which proved to be useful to confirm the diagnosis in 53% of cases with known disease. On the total sample of patients, the use of NGS technologies seems to be limited at this stage, but the results highlight the potential of the modern techniques of genome sequencing in the implementation of the diagnostic process; expanding the population of patients enrolled and the targeted application of these new technologies could allow wider knowledge on the phenotypic and genotypic profile of the considered diseases. Having a broader patient population, the creation of the expert system, of which the foundations were laid in the context of this project, will be possible through the collection of theoretical frequencies of clinical and diagnostic findings of some groups of rare diseases related (metabolic, mitochondrial), arising from the international literature and to be included in the system for processing of clinical scenarios. The expert system will allow the validation of diagnoses simulated by collecting new cases with known diagnosis, and the formulation of new diagnoses of known diseases or identification of new clinical entities, even with use of neural networks or Kohonen networks and Fuzzy methods. The methodology for collecting phenotypic and genotypic information in patients with rare diseases, proposed in this PhD project, proved sufficient to satisfy the requirements of applicability and suggested a number of future applications, which aim the expansion of knowledge on genomic variability of rare diseases and to reduce the number of patients without a diagnosis.

1 INTRODUZIONE

1.1 LE MALATTIE RARE: DEFINIZIONE

Malattia è qualsiasi alterazione dello stato di salute ed è considerata rara se si presenta nella popolazione generale con un tasso di prevalenza non superiore ad una data soglia, dipendente dalle legislazioni vigenti nel paese di riferimento, non essendo riconosciuta una definizione univoca ed esauriente a livello internazionale.

Secondo la definizione originariamente descritta su “Official Journal of the European Communities: EU legislation in Regulation (EC) N°141/2000 of the European Parliament of 16 december 1999”¹, ripresa nel “Community action programme on rare diseases 1999 – 2003”², primo programma inerente le malattie rare in ambito di salute pubblica europea, e quindi universalmente accettata dall’Unione Europea (UE), una patologia è definita rara se presenta una prevalenza di popolazione uguale o inferiore a 1/2000 abitanti nell’Unione Europea (0,05%) che in base alla popolazione generale europea, stabilisce un 6-8% di malati rari, cioè circa 30 milioni di individui. La medesima definizione è stata adottata dall’Agenzia Europea per i Farmaci (EMA) per la definizione di Farmaci Orfani.

A livello globale non esiste accordo sulla definizione di malattia rara; di fatto si tratta di definizioni di tipo statistico, stabilite per sostenere investimenti economici alle case farmaceutiche per ricerca e sperimentazione di nuovi farmaci in ambito di malattie rare, investimenti altrimenti non sostenibili a livello di mercato³.

Attraverso l’Orphan Drugs Act⁴, gli Stati Uniti sono stati il primo paese a fornire una definizione operativa di malattia rara, basata sulla soglia legislativa di 1/1250 abitanti, per un massimo 200.000 casi di malattia nella popolazione generale americana; secondo le stime dell’Organizzazione Nazionale per le Malattie Rare (NORD) e il National Institutes of Health (NIH), circa 25 milioni di americani soffrono di una malattia rara (prevalenza 8,4%)⁵.

In Giappone una malattia rara non può interessare più di 50.000 casi nella popolazione generale giapponese, con una soglia legislativa di 1/1250 abitanti⁶; a Taiwan⁷ una malattia è considerata rara se si presenta con una prevalenza non superiore a 1/10000 abitanti, così come in Australia⁸. Per alcune patologie considerate ultra-rare si

stimano tassi di prevalenza inferiori a 1/2.000.000 nella popolazione generale, o addirittura sono presenti solo casi sporadici o familiari.

La definizione di malattia rara, basata sulla prevalenza di popolazione generale, consente di includere un numero assai ampio di patologie a diversa eziologia, l'80% circa ad origine genetica (cromosomiche, monogeniche o complesse ad eziologia oligogenica, poligenica o multifattoriale, raramente epigenetica), alcune ad origine infettiva (ad esempio, morbillo materno-fetale), autoimmune (ad esempio, l'artrite cronica giovanile), tossica o tumorale. Ciascuna malattia interessa singolarmente un limitato numero di persone, ma dato il numero elevato di patologie, l'incidenza cumulativa è verosimilmente alta ed è possibile stimare che, globalmente, le malattie rare colpiscano una consistente quota della popolazione generale⁹.

Le Malattie Rare sono quindi un gruppo molto eterogeneo, sia per gli aspetti epidemiologici, sia sotto il profilo clinico. Tre quarti delle malattie rare può esordire alla nascita, in epoca perinatale o prima dei 2 anni d'età, un quarto appare dopo 40 anni; alcune possono avere manifestazioni cliniche meno gravi, ma circa l'80% ha serio impatto sulla spettanza di vita, nel 35% dei casi con exitus prima dell'anno di età. La disabilità correlata è sempre elevata e può compromettere gravemente la qualità di vita, in alcuni casi con perdita totale di autonomie personali. Si tratta pertanto di patologie spesso gravi, cronicamente debilitanti e invalidanti, frequentemente ad andamento progressivo e potenzialmente fatali; per tali ragioni rappresentano per molti paesi un grave problema di salute pubblica⁹.

1.2 PROFILO EPIDEMIOLOGICO DELLE MALATTIE RARE

È noto che le malattie rare determinano un elevato costo umano e sociale in termini di salute pubblica, per la complessità dei danni funzionali e della disabilità, e per la ridotta spettanza di vita ad esse correlate. Nonostante ciò, per motivi legati alla bassa prevalenza, sono state negli anni passati di scarso interesse sociale e scientifico, sono ancora oggi poco incluse nelle classificazioni internazionali e risultano poco visibili a livello d'informazione sanitaria; per molte di esse inoltre non sono ancora disponibili dati epidemiologici dalla letteratura scientifica¹⁰.

Negli Stati Uniti le malattie rare hanno raggiunto la ribalta dell'interesse sociale e scientifico a partire dal 1983, con l'adozione della Orphan Drug Act e l'istituzione di un Ufficio dedicato alle Malattie Rare presso il NIH^{4, 11}.

In Europa le malattie rare sono state ufficialmente definite una priorità in termini di salute pubblica e ricerca scientifica a partire dal 1999, con l'istituzione di una Task Force della Commissione Europea, che ha operato negli anni a programmi di politica sanitaria in tale ambito. Tale organo è stato successivamente sostituito dal Comitato di Esperti sulle Malattie Rare in Unione Europea (European Union Committee of Experts on Rare Diseases, (EUCERD)¹². Nonostante alcuni paesi non europei attuino oggi politiche in materia di farmaci orfani, l'Europa è attualmente l'unico paese che prevede specifiche politiche sanitarie a livello d'istituzioni pubbliche in ambito di malattie rare.

Nel 2006 uno studio ancora in corso, realizzato attraverso risorse derivanti dall'Organizzazione Europea per le Malattie Rare (Eurordis) e da Orphanet, con il supporto della CE, ha fornito le prime misure di prevalenza per alcune malattie rare, ricavate da una revisione completa della letteratura scientifica; dallo studio sono emerse drammaticamente la scarsità e la frammentazione delle misure epidemiologiche sulle malattie rare: per circa una quarantina di patologie non esistono dati epidemiologici in letteratura e per molte altre i dati disponibili sono scarsamente affidabili, di bassa qualità metodologica e derivanti da fonti poco coerenti o non confrontabili.

Gli studi epidemiologici implicano notevoli risorse in termini economici e la loro utilità in ambito di malattie rare non è stata ancora del tutto riconosciuta⁹; la raccolta d'indicatori epidemiologici di qualità risulta inoltre complessa e di non facile realizzazione, dato l'elevato numero di entità nosologiche conosciute e la difficoltosa rintracciabilità dei pazienti nei sistemi sanitari informativi pubblici, per ragioni altrettanto complesse legate all'eterogeneità dei sistemi classificativi di cui si vedrà in dettaglio in seguito¹⁰.

Sebbene la definizione di malattia rara si basi sulla prevalenza nella popolazione generale, molti studi epidemiologici forniscono misure di prevalenza alla nascita, applicando il termine non sempre in modo uniforme; in assenza di studi di follow-up per le malattie rare studiate risulta impossibile determinarne la prevalenza puntuale in altre epoche della vita, così come il tasso di mortalità calcolato sul numero di casi noti

di malattia. Nello stesso modo appare, ad esempio, difficile stimare la prevalenza di una malattia rara che colpisca solo il genere maschile (è il caso di tutte le patologie ad ereditarietà legata al cromosoma X), oppure uno specifico gruppo etnico, o una delimitata area geografica. In questi casi sarebbe fondamentale definire più correttamente la misura di prevalenza rispetto alla specifica popolazione riferimento⁹.

Pochi studi epidemiologici hanno fornito indicatori epidemiologici a livello di popolazione e si sono per lo più focalizzati su singole patologie o gruppi di patologie, nella grande maggioranza anomalie congenite, oppure rari disordini emorragici o ematologici tumori rari.

Orphanet, portale multilingue d'informazione sanitaria pubblica sulle malattie rare e i farmaci orfani, di cui si vedrà nel dettaglio in seguito, provvede al periodico aggiornamento di misure epidemiologiche, basato su ampie revisioni della letteratura scientifica e sul contributo di 588 tra database e registri di malattia. Il portale provvede inoltre a pubblicazione di misure epidemiologiche di prevalenza ed incidenza (o alla semplice notifica di casi isolati o familiari) di diverse centinaia di malattie rare, rappresentando in tal senso una valida risorsa per disporre di dati epidemiologici preliminari su alcune malattie rare e porre all'attenzione generale sulla necessità d'informazioni più estese ed affidabili per stimare l'impatto globale che queste malattie determinano¹⁰.

In tempi più recenti, grazie ai progressi raggiunti in ambito diagnostico e di tecnologie informatiche ed a seguito delle ultime raccomandazioni della CE sulla necessità di disporre di validi indicatori epidemiologici in ambito di malattie rare, iniziano ad essere disponibili le prime stime globali, derivanti da studi a livello di popolazione.

In Italia la prevalenza complessiva delle malattie rare, stimata sulla base dei dati del Registro Interregionale Piemonte e Valle d'Aosta (21.062 casi nel periodo giugno 2005-2013), risulta pari a 4,5/1.000 abitanti¹³.

Secondo un altro studio con ampio campione di popolazione, derivante dal Registro Regionale per le Malattie Rare del Veneto¹⁴, le condizioni rare considerate presentano una prevalenza globale pari a 3.3/1000 abitanti, che varia dallo 0.003/1000 per le patologie connesse all'epoca perinatale, allo 0.54/1000 per le anomalie congenite.

Un terzo dei pazienti affetti è in età pediatrica e i gruppi di malattie con percentuale di prevalenza maggiore in questa fascia d'età riguardano le malformazioni congenite (43%), le patologie ematologiche (15.5%), i disordini metabolici ereditari (8.7%) e i tumori rari (7.2%). Le anomalie congenite sono il gruppo di patologie a più elevata percentuale di mortalità (35.6%), seguite dai disordini metabolici ereditari (20.6%). In età adulta il gruppo di patologie a percentuale di prevalenza maggiore sono le anomalie di occhio ed annessi (17.6%), con una prevalenza globale di 0.44/1000, seguite dalle patologie ematologiche (15.7%) e dalle malattie del sistema nervoso centrale (13.0%), che sono anche responsabili della maggior quota di mortalità (53.4%). La prevalenza globale delle malattie rare stimata in questo studio è pari a 6.6/1000 individui e può aumentare fino a 19.5/1000 se si considerano i tumori rari, che presentano una prevalenza variabile tra 1.77/1000 per carcinomi del sistema endocrino e lo 0,0003/1000 per tumori trofoblastici placentari; secondo queste stime tra i 6.500.000 ed i 9.880.000 di persone in Europa sono malati rari, ovvero tra l'1.3-2% della popolazione generale.

In Francia la prevalenza cumulativa delle malattie rare è stimata tra il 4% e il 6%, con la maggioranza delle 500 malattie rare più frequenti con prevalenza inferiore a 1/100 000. Il tasso d'incidenza cumulativa annuale è tra l'1,3% e il 2,2% della popolazione generale per le 199 patologie più comuni¹⁵.

Un altro studio francese riporta una prevalenza globale di malattie rare nei bambini di 7 anni d'età, stimata al 2,1/1000 in base ai dati del "Registre des Handicaps de l'Enfant et Observatoire Perinatal (RHEOP)", registro per pazienti pediatriche con disabilità e per la sorveglianza perinatale¹⁶.

In ambito non europeo, uno studio di popolazione condotto a Taiwan fornisce un profilo epidemiologico della popolazione di malati rari sulla base dei dati del Registro Nazionale per la disabilità nel periodo 2002-2011 e stima la prevalenza globale delle malattie rare nella popolazione taiwanese tra 0,02-0,74/10.000⁷.

Negli USA è stato condotto uno studio di popolazione su adulti, con dati derivanti dal "Cancer in North America (CINA) dataset" tra il 1995 e il 2004, che riguardano circa l'80% della popolazione statunitense. Secondo questo studio, che rispetta la definizione di tumore raro in base ad una prevalenza di popolazione non superiore a 15 casi/100.000 individui/anno, i tumori rari rappresentano circa il 25%

delle neoplasie dell'età adulta, in USA con una prevalenza globale stimata pari a 6,130.25/1,000,000/anno¹⁷.

Queste misure di prevalenza risultano molto variabili, poco confrontabili e con buona probabilità sottostimate, non considerando la quota d'individui affetti ancora senza diagnosi, che non rientrano nei registri; si vedrà in seguito come trascuratezze, errori e ritardi diagnostici in ambito di malattie rare siano un problema frequente. Le stime ufficiali accettate in Europa, in base alle quali tra il 6-8% della popolazione generale europea è affetta da malattie rare⁴, ovvero circa 30 milioni di individui, possono ragionevolmente essere considerare poco coincidenti ad alcuni dati di realtà che iniziano ad emergere in ambito epidemiologico.

Da queste premesse sono evidenti la complessità del profilo epidemiologico delle malattie rare e le attuali ancora scarse conoscenze sull'impatto che esse determinano nella loro globalità (e non come singole patologie o gruppi di patologie) sulla popolazione generale. Appare emergente la necessità di ampliare le conoscenze sul profilo epidemiologico malattie rare e definirne l'impatto sia in termini di morbilità e mortalità delle malattie rare, sia rispetto a misure epidemiologiche di altre più comuni condizioni patologiche, lesioni o fattori di rischio. Queste conoscenze sono ancor oggi difficili da stimare, sulla base delle risorse epidemiologiche disponibili, ma sono necessarie per adeguare le politiche sanitarie ai reali bisogni della popolazione e per guidare le decisioni ed i finanziamenti pubblici. In tal senso la produzione di misure epidemiologiche attendibili a livello globale è una tendenza emergente nella comunità scientifica e, in ambito di politiche socio-sanitarie, largamente sostenuta nelle raccomandazioni degli organi competenti della CE. Le specifiche caratteristiche delle malattie rare, come la bassa prevalenza nella popolazione generale, il numero elevato di entità nosologiche diverse (patologie e sindromi), la generale scarsità di conoscenza e di competenze in materia, la difficoltà nel processo diagnostico, il numero elevato di malati rari senza diagnosi e, per la maggior parte, la mancanza di terapie efficaci rendono quindi ragione di specifiche problematiche connesse in termini di politiche socio-sanitarie e della necessità di strategie congiunte per la diagnosi precoce, la cura e la prevenzione dell'elevata morbilità, della mortalità perinatale o precoce e della notevole riduzione della qualità di vita degli individui affetti.

1.3 SISTEMI CLASSIFICATIVI PER LE MALATTIE RARE

La ragione principale della complessità del profilo epidemiologico delle malattie rare, spesso in mancanza di studi di buona qualità metodologica, sembra essere stata l'assenza di una classificazione univoca a livello globale e di un sistema di corretta codifica per le malattie rare, che ne consentano la rintracciabilità all'interno dei sistemi informativi sanitari³. Il numero di malattie rare note, già molto ampio, è in costante aumento, compreso attualmente tra 5000 e 8000 entità, incluse singole malattie, anomalie morfologiche o biochimiche e quadri sindromici, nell'80% dei casi legate ad una causa genetica¹⁸, quasi sempre ad ereditarietà mendeliana e spesso rientranti in diverse classificazioni fenotipiche e/o genotipiche non completamente sovrapponibili. Più in generale infatti i sistemi di raccolta dell'informazione fenotipica e genotipica delle malattie sono ad oggi ampiamente eterogenei, privi di terminologie e nomenclature standard, caratterizzati da differenti metodologie ed ontologie di riferimento, scarsamente efficienti in termini di inter-operabilità e poco utili, nella loro forma attuale, alla raccolta dell'informazione completa in ambito di umane ad ereditarietà mendeliana¹⁴. La prospettiva ideale a livello globale sarebbe poter disporre di sistemi di raccolta dell'informazione sulle malattie ereditarie, comuni o rare, nelle sue caratteristiche attuali in epoca di sequenziamento genomico¹⁹, cioè nella sua completa corrispondenza fenotipo-genotipica, con lo sviluppo di database di varianti genomiche in grado di classificare le malattie ereditarie secondo la variabilità genomica e del corrispondente profilo fenotipico e quindi generare nuova conoscenza sull'impatto di tale variabilità sulla salute umana²⁰. Attualmente le problematiche legate all'analisi computazionale incrociata e ad una classificazione globale di tutti i fenotipi e delle varianti genomiche associate sono innumerevoli e non esistono linee guida ufficiali complete ed esaurienti in merito a metodologia, terminologia di riferimento etc²¹.

In ambito di malattie rare il superamento di queste criticità ed il raggiungimento di un consenso e di una classificazione nosologica fenotipo-genotipica di riferimento è una necessità emergente per rintracciare queste malattie nei sistemi sanitari informativi, predisporre strumenti utili a generare nuova conoscenza sul loro profilo epidemiologico e sull'impatto a livello di popolazione in termini di morbilità e di mortalità, e quindi per rendere operativi, in ambito di salute pubblica, piani idonei a garantire ai pazienti affetti l'accesso a servizi diagnostico-terapeutici di alta qualità.

1.3.1 Descrizione del fenotipo

Le malattie umane ad ereditarietà mendeliana note, rare o comuni, sono al momento svariate migliaia, ognuna con una specifica e distinta combinazione di caratteristiche fenotipiche; esistono un'ampia varietà di sistemi descrittivi e classificativi, alcuni unicamente dedicati alla classificazione del fenotipo delle malattie umane e dotati di specifiche terminologie di riferimento, altri con caratteristiche di database genomici contenenti ampie informazioni sui profili fenotipici variante-correlati.

Per le malattie rare, nella maggior parte dei paesi, i sistemi sanitari informativi utilizzano ancora la classificazione internazionale delle malattie, International Classification of Diseases (ICD) nelle versioni ICD-10 ed ICD-9, un sistema classificativo di tipo clinico, unidimensionale, basato sulle caratteristiche fenotipiche delle malattie, che presenta tuttavia alcuni limiti di codifica e classificazione delle malattie più rare; l'ICD-10 ad esempio include solo 500 circa malattie rare e ne identifica un codice specifico solo nella metà dei casi. Per tale ragione non rappresenta lo strumento più affidabile per l'identificazione dei pazienti affetti all'interno dei sistemi informativi sanitari pubblici e per la registrazione degli stessi in banche dati nazionali o internazionali¹⁴. Tali criticità sono state riconosciute già a partire dal 2009 dalla CE, che nelle raccomandazioni del Consiglio ha identificato come prioritarie la codifica e la rintracciabilità delle malattie rare nei sistemi informativi d'assistenza sanitaria, sostenendo l'implementazione dei sistemi classificativi attualmente in uso (WHO International Classification of Diseases-ICD), per disporre di adeguati indicatori epidemiologici e quindi attuare, in ambito di salute pubblica, strategie idonee a garantire ai pazienti affetti l'accesso a servizi diagnostico-terapeutici di alta qualità²².

Dal 2007, in accordo alle direttive dell'OMS sull'aggiornamento e la revisione della classificazione ICD-10 e sulla stesura del nuovo manuale ICD-11, sono stati istituiti dei Topic Advisory Groups (TAG), organi consultivi con funzione di pianificazione e coordinamento dell'intero processo di revisione per aree specifiche del sistema ICD²². Il TAG specifico per le malattie rare ha individuato Orphanet come risorsa principale per la raccolta d'informazioni preliminari, necessarie all'aggiornamento di quest'area chiave del sistema classificativo ICD, nella prospettiva

di utilizzare la terminologia ed il sistema di classificazione Orphanet come modello per la stesura della nuova classificazione ICD-11. Oggi, grazie al lavoro del gruppo TAG specifico per le malattie rare è possibile la mappatura uno ad uno dei codici Orpha con i termini specifici ICD-11; attualmente i codici Orpha stanno entrando in uso nei sistemi informativi ospedalieri e si prevede che nei prossimi anni aumenterà significativamente la rintracciabilità e la visibilità delle malattie rare all'interno dell'intero sistema informativo sanitario.

La nomenclatura Orphanet si posiziona ad un crocevia di ontologie differenti e contribuisce ad una medicina con caratteristiche d'inter-operabilità e trasversalità; si tratta di un processo dinamico in continuo aggiornamento, dotato tuttavia già di caratteristiche di stabilità e coerenza sufficienti a configurare il sistema classificativo Orphanet come potenziale terminologia di riferimento internazionale per la malattie rare¹⁰.

Orphanet, portale multilingue d'informazione sanitaria pubblica sulle malattie rare e i farmaci orfani, nasce nel 1997 per iniziativa dell'Istituto Nazionale Francese per la Salute e la Ricerca Medica (INSERM) e della Direzione Generale della Sanità Francese ed è stato finanziato come progetto europeo nel 2000 dalla Commissione Europea (UE). Le informazioni in esso contenute sono di facile e libero accesso attraverso il sito web (www.orpha.net), la piattaforma di download dedicata (www.orphadata.org) e derivano da fonti molto eterogenee dal punto di vista strutturale e metodologico in riferimento alla raccolta di dati epidemiologico-clinici.

Fino al 2007 Orphanet era costituito da un semplice, ma esaustivo elenco di malattie rare, utile principalmente per informazioni inerenti i prodotti farmaceutici, creati per il trattamento di patologie singole o in limitato gruppo. La nomenclatura Orphanet, mappata in base alla terminologia ICD-10 seppur con scarso livello qualitativo, è stata utilizzata in passato per codifica delle cartelle cliniche ospedaliere con risultati promettenti: nel 2006 in Francia è stato creato CEMARA un database d'informazioni cliniche di pazienti provenienti da 55 centri esperti in rete, che ha consentito il recupero di un'enorme quantità d'informazioni d'interesse medico scientifico, con l'identificazione dei pazienti affetti da malattia rara nel sistema

informativo sanitario e, in alcuni casi, con la codifica di codici specifici di malattia per malati rari senza diagnosi precisa²³.

Dall'aprile 2007, in linea con le direttive del TAG specifico per le malattie rare, è attivo nella CE un programma operativo che presiede all'intero processo di riclassificazione delle malattie rare attraverso Orphanet e che ha identificato come priorità la codifica delle entità nosologiche ad un più alto grado di precisione e completezza, ed una classificazione che ne rispetti la multidimensionalità fenotipo-genotipica e ne migliori la rintracciabilità nei sistemi sanitari informativi, soprattutto in termini di mortalità e morbilità²⁴.

L'obiettivo del processo di aggiornamento del sistema Orphanet è stato quello di ampliare al più alto grado la conoscenza in ambito di malattie rare, a beneficio di ampie categorie di utenti, a partire dai revisori dei gruppi di lavoro TAG, passando da pazienti e famiglie fino ad arrivare a medici specialisti, clinici, ricercatori, manager e operatori sanitari, bioinformatici che, attraverso pubblicazioni scientifiche e lavoro in rete tra gruppi di esperti in circa 40 diversi paesi, ancora collaborano alla raccolta e all'archiviazione dati.

La quarta versione di Orphanet, uscita nel 2008 non è più un semplice inventario di malattie, ma diviene un database relazionale, la cui logica parte dal concetto centrale di entità nosologica unica e distinta, codificata ed inserita in una sistema classificativo multigerarchico con struttura ad albero, a infiniti livelli di granularità. Il database integra la classificazione delle malattie rare, le relazioni tra differenti classi principali di oggetti (patologie o gruppi di patologie, manifestazioni cliniche o processi patogenetici, geni, dati epidemiologici, informazioni di base su ereditarietà e modalità di trasmissione) ed i collegamenti a molteplici fonti esterne, quali sistemi classificativi o i database clinici e database genomici, liste di farmaci orfani e di prodotti affini, o di altri servizi inerenti l'ambito delle malattie rare (centri clinici e di ricerca, laboratori, associazioni di pazienti), informazioni enciclopediche, o dettagli più tecnici quali linee guida alla pratica clinica, o articoli scientifici dalla letteratura internazionale.

La piattaforma Orphadata, lanciata nel 2011, garantisce l'integrazione e l'elaborazione dell'enorme quantità di dati che via via popolano il database, rendendo disponibile agli utenti, per consultazione libera o regolamentata, l'estrazione di una

varietà d'informazioni complesse sotto il profilo clinico-epidemiologico, favorendo in maniera sistematica il flusso di conoscenza in ambito di malattie rare, attraverso i sistemi informativi sanitari.

Il database è mantenuto in continuo aggiornamento attraverso la gestione incrociata di dati provenienti da diverse fonti complementari di conoscenza e la complessità delle informazioni derivanti determina la necessità di operare secondo una logica d'inter-operabilità semantica tra differenti domini di conoscenza, o ontologie, e le relative terminologie di riferimento, attraverso una mappatura che consenta la relazione corretta tra termini diversi concettualmente corrispondenti.

La sezione dedicata all'ontologia delle malattie rare (Orphanet Rare Disease Ontology-ORDO) è stata creata in collaborazione con l'European Bioinformatics Institute (EBI) ed una serie di risorse esterne, tra cui i National Genetic Reference Laboratories (Manchester, UK) e l'International Union of Basic and Clinical Pharmacology (IUPHAR), anche con l'obiettivo di facilitare l'interoperabilità semantica con altre ontologie, assicurando la coerenza ed il rispetto delle regole logiche tra i diversi modelli di conoscenza.²⁵

A tal proposito a partire dal 2012 Orphanet ha avviato la mappatura della propria terminologia secondo il sistema Unified Medical Language System (UMLS), in particolare con le terminologie SNOMED CT, MeSH e MedDRA¹⁰.

UMLS è un compendio di nomenclature standardizzate di termini scientifici biomedici, progettato e creato nel 1986 dalla US National Library of Medicine (NLM), attualmente aggiornato ogni tre mesi ed utilizzabile gratuitamente. Esso presenta un modello di mappatura incrociata tra diverse terminologie, che consente l'interoperabilità dominio di conoscenza; può essere visto anche come un'ontologia concetti biomedici ed è destinato all'utilizzo per sviluppo ed implementazione di sistemi bio-informatici.

L'obiettivo di UMLS è migliorare l'accesso alla letteratura scientifica e il recupero di fonti, facilitando lo sviluppo di sistemi informatici che traducano il linguaggio biomedico, superando le difficoltà legate alla variabilità dei termini utilizzati per descrivere medesimi concetti ed alla dispersione di informazioni utili in numerosi differenti database e sistemi classificativi eterogenei.

Il database UMLS comprende oltre 1 milione di concetti biomedici e 5 milioni di termini specifici per tali concetti, derivanti da oltre 100 vocabolari standardizzati ed utilizzati in diversi sistemi di classificazione²¹: alcuni esempi ICD-10, MeSH, SNOMED CT e OMIM.

Date le dimensioni e la complessità di UMLS sono presenti errori di ambiguità e ridondanza dei termini, rapporti gerarchici non del tutto chiari, relazioni semantiche tra termini e concetti non coerenti. UMLS provvede manualmente al controllo di qualità con notevole spesa in termini di tempo e costi, attualmente con sistemi e strumenti automatici in via di sviluppo²⁶.

Systematized Nomenclature of Medicine, Clinical Term (SNOMED CT) è un sofisticato vocabolario biomedico, pubblicato per la prima volta nel 2002 allo scopo di fornire una nomenclatura di riferimento internazionale per le cartelle ospedaliere informatizzate, potenzialmente utile alla raccolta di dati clinici standardizzati a scopo epidemiologico ed assistenziale²⁷. La maggior parte degli studi presenti in letteratura inerenti la nomenclatura SNOMED-CT si basano ancor oggi su modelli teoretici o pre-clinici, mentre sono pochi gli studi d'implementazione. La necessità emergente a livello scientifico sembra essere quella di armonizzare la terminologia SNOMED-CT con altre nomenclature standardizzate per superare le costanti problematiche d'interoperabilità semantica delle cartelle cliniche informatizzate in uso nei diversi paesi, legate soprattutto alla grande ambiguità, granularità e variabilità dei termini utilizzati o addirittura del testo libero. A tal scopo la terminologia SNOMED-CT è da sempre sottoposta a revisione semestrale e ad oggi è ormai considerata relativamente completa, idonea all'utilizzo in ambito clinico, rappresentando la terminologia di riferimento in uso nelle cartelle cliniche informatizzate in circa 19 diversi paesi²⁸; le malattie rare sono tuttavia ancora poco rappresentate all'interno della nomenclatura SNOMED-CT e la mappatura incrociata delle terminologie Orphanet e SNOMED-CT può consentire l'integrazione dei termini mancanti, con notevole beneficio in termini d'usabilità²⁹.

Medical Subject Headings (MeSH) è un esteso thesaurus di termini biomedici creato e mantenuto anch'esso grazie alla NLM degli Stati Uniti, allo scopo di indicizzare gli articoli scientifici delle oltre 5000 riviste del database bibliografico PubMed/Medline e nel catalogo di libri della NLM³⁰.

MesDRA è un dizionario terminologico medico standardizzato, altamente specifico sviluppato negli ultimi anni 90 nell'ambito dell'International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) con il fine di facilitare l'integrazione informazioni per la regolamentazione internazionale dei prodotti medicali ad uso umano.

L'allineamento della nomenclatura Orphanet con queste terminologie sta rendendo disponibili collegamenti a nuove "queries" specifiche di PubMed, favorendo anche in questo caso una miglior rappresentazione delle malattie rare all'interno della NLM.

All'interno del sistema classificativo di Orphanet, che rispetta appunto una struttura multigerarchica ad albero con infiniti livelli di granularità, a seconda dell'ontologia o dell'area semantica d'interesse ogni malattia (o gruppo di malattie) può essere codificata e classificata in base a differenti percorsi logici, che prendono forma da un continuum fenomenico a completezza crescente. Ogni differente percorso logico, sia esso clinico, fisiopatologico o anatomico, terapeutico, eziologico, genomico etc., rappresenta una ramificazione principale o secondaria della struttura ad albero, dove ciascun nodo rappresenta un'entità nosologica (patologia, gruppo di patologie e/o sottotipo di patologia). L'entità mapperà a differenti livelli della ramificazione (clinico a livelli più terminali o alti ad esempio e genomico a livelli più apicali o profondi) e potrà pertanto essere rappresentata all'interno di più ramificazioni, a seconda dei differenti percorsi logici sottostanti (parentela multipla), restando sempre rintracciabile all'interno della classificazione stessa. L'altra caratteristica cardine del database relazionale è infatti la rintracciabilità delle singole entità nosologiche, che rappresentano il concetto centrale dell'intera logica del database; tale caratteristica viene costantemente garantita dal numero Orpha, codice classificativo specifico ed univoco, che, una volta assegnato, si riferisce sempre alla medesima entità e rimane costante nel tempo anche a fronte della continua evoluzione del database stesso, con codifica di nuove entità nosologie, riclassificazione delle entità nosologie note, aggiornamento della terminologia ed ampliamento della sinonimia.¹⁰

L'"Online Mendelian Inheritance in Man" (OMIM) nacque nel 1966 ad opera di Victor McKusick, nella prima edizione intitolata "Mendelian Inheritance in Man"

(MIM), come un semplice compendio della letteratura scientifica su malattie umane ad ereditarietà mendeliana. Il compendio è stato poi rivisto ed aggiornato fino al 1998 in 12 riedizioni periodicamente pubblicate e curate manualmente, ma già a partire dai primi anni '90, con l'avvento del sequenziamento genomico si è strutturato come un database, ovvero un articolato catalogo di geni e varianti genetiche, correlate ad alcuni quadri fenotipici ereditari; ad ogni entità è assegnato un numero identificativo MIM, univoco ed immutabile nel tempo; i numeri MIM sono ampiamente utilizzati in letteratura e hanno dimostrato la loro utilità come codici identificativi di riferimento nella classificazione della variabilità fenotipica umana ³¹.

La distribuzione e lo sviluppo dei software per il mantenimento di OMIM sono garantiti dal National Center of Biotechnology Information (NCBI) presso la NLM, che dal 1995 ne garantisce anche consultazione gratuita on-line (<http://www.ncbi.nlm.nih.gov/omim/>) ³².

Negli ultimi 40 anni, OMIM ha catalogato le malattie umane ad ereditarietà mendeliana concentrandosi sulle relazioni causative e le corrispondenze tra geni e varianti genomiche e profili fenotipi. Nel novembre 2010, OMIM conteneva oltre 20.267 voci relative a 13.606 geni e oltre 7.000 quadri fenotipi. OMIM viene mantenuto in continuo aggiornamento, al passo con le scoperte generate dal sequenziamento genomico, con l'obiettivo prioritario di catalogare al miglior dettaglio i dati relativi alle varianti genomiche con rilevanza clinica o significato funzionale³³.

Tra i limiti di OMIM è da annoverare l'impossibilità di condurre un data mining sistematico, a causa dell'utilizzo di una terminologia non standardizzata all'interno del database. Vari tentativi sono già stati condotti in tal senso, allineando le nomenclature MeSH o ULMS alla terminologia OMIM; queste analisi hanno dato risultati promettenti, ma sono state riscontrate numerose difficoltà dovute alle caratteristiche dei sistemi di nomenclatura utilizzati, che non sono stati progettati in modo specifico per la descrizione di quadri fenotipi umani e catalogazione delle malattie ereditarie. ULMS ad esempio consente la mappatura solo di circa il 34% dei fenotipi specifici di malattia associati ad una data variante genetica ³⁴.

Human Phenotype Ontology (HPO) fornisce una terminologia standardizzata basata su una rappresentazione gerarchica dell'anatomia e dei fenotipi umani ed è stato

sviluppato a partire dalle entità contenute in OMIM, con l'obiettivo di catalogare tutte le anomalie fenotipiche comunemente presenti nelle malattie umane ad ereditarietà mendeliana. HPO sembra vantaggioso perché dotato di una terminologia basata sul sapere medico piuttosto che su ontologie di text-mining. L'ontologia HPO è un modello di organizzazione dei dati che rappresenta concetti, relazioni ed attributi con una maggiore flessibilità e potenza descrittiva, rispetto ad altre serie di dati utilizzabili per l'analisi fenotipica, per la possibilità di stabilire relazioni concettuali uno a molti tra termini, all'interno di una struttura gerarchica, che consente la descrizione, la ricerca dati ed il ragionamento logico sui termini utilizzabili a diversi livelli di granularità e completezza. Grazie alla sua struttura gerarchica HPO è in grado di gestire termini a significato ambiguo, sono spesso utilizzati in letteratura scientifica ²¹.

Oltre ai sistemi classificativi, alle nomenclature ed alle ontologie descritte per la classificazione dei quadri fenotipici delle malattie umane ereditarie, esiste una serie di database di varanti genomiche, contenenti ampie sezioni di dati riguardanti le caratteristiche fenotipiche associate a varianti genomiche, tra cui ClinVar ^{30,35}, PharmGKB ³⁶, Leiden Open Variation Database (LOVD), Alzheimer Disease and Frontotemporal Dementia Mutation (AD and FDM), Immunodeficiency database (IDBases), Genotypes and Phenotypes (dbGaP), PhenoDB etc. In particolare PhenoDB ³⁷ è un portale Web protetto per la raccolta, la conservazione e l'analisi di caratteristiche fenotipiche, creato da Hamosh e collaboratori presso la John Hopkins University ³⁸. Il database disponibile gratuitamente (<http://phenodb.net/downloads>) riporta caratteristiche fenotipiche secondo una struttura gerarchica, basata sul modello adottato da OMIM per le sinossi cliniche. Il database è una risorsa utile e robusta per la raccolta, l'archiviazione e l'analisi dei dati fenotipici, soprattutto nel contesto delle tecnologie di sequenziamento del genoma umano e delle strategie volte le varianti genomiche responsabili delle malattie.

Alcuni di questi database, come si vedrà meglio in seguito, sono già operativi, altri sono in via di sviluppo ed implementazione, di pari passo con la creazione di strumenti informatici che consentano l'interrogazione sistematica dei dati disponibili nei differenti database, la raccolta, la conservazione e l'analisi delle caratteristiche fenotipiche e delle informazioni cliniche utili all'analisi del genoma umano.

1.3.2 Descrizione del genotipo

In contrapposizione alla cosiddetta “first generation technology”, ovvero la tecnica di sequenziamento di singoli geni (o di un ristretto gruppo di geni) del DNA tramite la metodica dell'elettroforesi capillare, ideata da Sanger e colleghi³⁹, ed i cui continui sviluppi tecnologici, dal 1970 in poi, hanno portato al primo sequenziamento completo del genoma umano nel 2004, nello stesso anno viene commercializzata la tecnologia Next Generation Sequencing (NGS) come tecnica di sequenziamento genomico di nuova generazione, cosiddetta “second generation sequencing”; tra il 2005 ed il 2010 compare sul mercato una serie di prodotti tecnologici NGS sempre più avanzati ed efficienti, in grado di generare sequenze di lettura molto lunghe⁴⁰.

Le tecnologie NGS hanno, negli ultimi anni garantito alla comunità medico scientifica la possibilità di studiare una vasta gamma di malattie genetiche³³, favorendo la creazione di un flusso di lavoro uniforme in ambito di ricerca clinica, con crollo dei costi e notevole riduzione dei tempi d'analisi dei dati genomici, rispetto alle tecniche molecolari tradizionali⁴¹. Per tale ragione si è assistito, a livello mondiale, alla nascita di centri e progetti di ricerca altamente specializzati in tecniche di sequenziamento genomico ed allo sviluppo di database di varianti genomiche pubblicamente accessibili, di cui si vedrà nel dettaglio seguito³⁸. Alcuni esempi di progetti di ricerca sono l'”Undiagnosed Diseases Program” (UDP) o i progetti dei “Center of Mendelian Genomics (CMGs) presso la University of Washington, la Yale University e la John Hopkins University⁴². La collaborazione con risorse internazionali, quali l'International Rare Disease Research Consortium, il Finding of Rare Disease Genes from Canada e i Centers from East Asia and Europe, ha inoltre favorito il conseguimento di obiettivi utili sia per la ricerca clinica, sia in ambito assistenziale, attraverso l'implementazione delle applicazioni delle tecnologie NGS e la condivisione di metodologie e dati genomici a livello globale. Il continuo accrescimento della conoscenza sulle malattie ad ereditarietà mendeliana, con possibilità di esplorare i sistemi biologici e di comprendere le basi eziopatogenetiche di specifiche malattie ad un dettaglio mai raggiunto prima, ha posto le basi per lo sviluppo di nuovi trattamenti. Le tecnologie informatiche hanno giocato un ruolo chiave nel passaggio dalla semplice identificazione, all'interpretazione del ruolo patogenetico delle sequenze varianti del genoma umano, migliorando l'efficienza e l'accuratezza del lavoro dei clinici e dei ricercatori e negli ultimi 4-5 anni

sono stati sviluppati algoritmi e procedure analitiche per la standardizzazione e per l'interpretazione univoca dei dati delle sequenze di lettura in ambito medico. Con l'evoluzione dei sistemi tecnologici si è sviluppato un numero crescente di metodi preparazione dei campioni e di strumenti di analisi dei dati con una conseguente immensa varietà di applicazioni scientifiche.

Le applicazioni delle tecnologie NGS in ambito di ricerca clinica sono conosciute come “Whole Genome Sequencing” (WGS) e “Whole exome sequencing” (WES), rispettivamente per lo studio dell'intero genoma umano e dell'esoma umano, ovvero la parte codificante del genoma, pari a circa l'1% del genoma stesso e contenente fino all'85% delle mutazioni ad effetto patogeno fin qui conosciute⁴³.

In epoca attuale a livello mondiale si stima che, grazie a queste tecniche, siano identificate circa 5 nuove entità nosologiche alla settimana, con continua scoperta di nuovi e differenti profili genotipici; tra il 2007 ed il 2014 il numero di quadri fenotipi legati ad una variante genomica nota è più che raddoppiato, con 428 nuove entità nosologie aggiunte solo tra il 2013 ed il 2014. Inoltre l'interesse generale per queste tecnologie sta crescendo di pari passo con la conoscenza in ambito di variabilità genomica umana ed è ormai evidente che la crescente diffusione ha rivoluzionato lo studio delle malattie rare. Nel particolare ambito di ricerca sulle malattie rare la WES si è largamente dimostrata la tecnica più efficiente, in termini di riduzione di tempi e costi, per la raccolta e l'analisi dei dati genomici. Il numero di varianti genomiche coinvolte in malattie rare ad ereditarietà mendeliana è in continuo aumento, ad oggi con identificazione di oltre 300 regioni esoniche per circa 150 geni causativi tramite tecnologia WES⁴³.

La ricerca clinica che si avvale di tecnologia NGS è basata su sistemi e strategie bio-informatiche per l'identificazione di varianti genomiche rare su campioni di piccole dimensioni. Tecnologie NGS sono già stata applicate con successo in complesse indagini genetiche, ad esempio in ambito di genomica del cancro o per l'analisi di frequenza di mutazione a livello di popolazione. La Sindrome di Kabuki e la Sindrome di Miller sono state due delle prime malattie rare la cui eziologia è stata chiarita grazie alle tecnologie di WES. Nel 2009 Ng e collaboratori⁴⁴ hanno utilizzato per la prima volta con successo la tecnologia NGS per l'identificazione di una patologia ad

ereditarietà mendeliana a causa sconosciuta, la Freeman-Sheldon Syndrome in 12 individui ¹⁹. In anni successivi sono stati descritti numerosi geni responsabili di patologie ad ereditarietà mendeliana, attraverso avanzate tecniche d'ibridazione per identificazione e sequenziamento dell'intero genoma umano. Studi incrociati sono stati condotti per identificare rare sequenze genomiche varianti a probabile significato patogenetico, in relazione a quadri fenotipici d'interesse e la NGS è stata applicata con successo anche nell'identificazione di mutazioni de novo in casi sporadici di ritardo mentale. Altri studi di analisi linkage combinata alla NGS in membri di famiglie affette o in individui affetti da malattie autosomiche recessive o X-linked, ne hanno identificato le regioni esoniche a significato patogenetico.

Il gruppo di lavoro e ricerca che nel 2013 ha portato alla creazione di PhenoDB, di cui si è visto a proposito dei sistemi di raccolta dell'informazione fenotipica, ha sviluppato una recente iniziativa, nell'ambito dei progetti CMG, per l'identificazione e la mappatura dei geni causativi le malattie mendeliane a causa ancora sconosciuta, tramite l'utilizzo di dati derivati da tecnologie WES ³⁷.

Presso la Duke University della Carolina del Nord statunitense è stato istituito il "Center for Human Genome Variation", che dal 2005 è responsabile di numerosi studi che hanno utilizzato la tecnologia NGS per l'identificazione di regioni genomiche responsabili di alcune specifiche malattie, con importanti implicazioni diagnostico-assistenziali.

Il gruppo di ricercatori del Cold Spring Harbor Laboratory ha condotto studi sulla variabilità genetica in ambito di severi disordini neuropsichiatrici, con particolare interesse per alcune malattie rare, tra cui autismo, schizofrenia e sindrome di Tourette.

Tecnologie di sequenziamento genomico NGS sono state utilizzate anche in limitati ambiti clinici, come tecniche diagnostiche supplementari, guidando con successo le decisioni mediche nel corso del percorso diagnostico e nel prossimo futuro sono attesi nuovi e sempre più numerosi dati genotipici relativi alle 3,750 patologie mendeliane note e ad altri disordini mono o poligenici di cui l'eziologia risulta ancora del tutto o in parte sconosciuta. È prevedibile inoltre un'ulteriore riduzione dei costi e lo sviluppo di nuovi metodi standardizzati che le renderanno applicazioni routinarie anche in ambito assistenziale diagnostico. La tecnologia WES può ad esempio essere utile in

quei casi in cui manchi la diagnosi clinica precisa, ma vi sia un sospetto diagnostico rispetto ad una regione genomica specifica potenzialmente causativa o in casi di malattie geneticamente eterogenee, i cui soggetti affetti rimangano senza conferma diagnostica anche dopo esecuzione di test molecolari tradizionali.

In generale le tecnologie NGS si sono dimostrate risolutive soprattutto nell'identificazione di varianti genomiche in relazione a profili fenotipi molto precisi, caratterizzati dalla combinazione di determinate caratteristiche cliniche che orientavano verso a specifiche malattie; è questo il caso ad esempio delle malattie metaboliche, neurologiche e degli organi di senso (ipofisine o ipoacusia ereditarie).

Ad oggi centinaia, forse migliaia di pubblicazioni scientifiche hanno identificato le tecnologie NGS, in particolare la WES, come risorsa nuova ed efficace nello studio di singoli geni, dei tumori, dell'evoluzione umana e della variabilità genomica delle popolazioni.

Senza dubbio le tecnologie NGS o la cosiddetta epoca “second generation sequencing” hanno permesso il definitivo passaggio dalla genetica clinica alla medicina genomica, con potenzialità immense per l'accrescimento della conoscenza globale sui processi patogenetici delle malattie umane ad ereditarietà mendeliana. I progressi raggiunti rappresentano tuttavia solo l'inizio di una serie di sforzi coordinati della comunità scientifica i cui obiettivi restano l'implementazione del processo diagnostico e del percorso assistenziale e terapeutico del paziente affetto. L'applicazione di queste tecnologie per la diagnosi clinica rimane difficile, soprattutto in ambito di malattie rare ad eziologia complessa o multifattoriale, principalmente a causa della difficoltà d'interpretazione delle sequenze genomiche varianti, in relazione a specifiche caratteristiche fenotipiche.

Negli ultimi dieci anni l'avvento delle moderne tecnologie NGS, la bio-ontologia e il “Semantic Web”, così come il ruolo crescente dell'Information Technology (IT) in ambito ospedaliero hanno reso disponibile, come visto, una grande e crescente varietà di informazioni sulle variabilità genomiche tra gli individui, intese come sequenze varianti di DNA o mutate rispetto alla sequenza di riferimento standard; tali informazioni sono spesso di difficile interpretazione risultando, nella maggior parte dei casi, ancora impossibile da determinare la relazione tra l'informazione genotipica ed il

fenotipo e, quindi, il reale impatto delle varianti genomiche in termini di salute umana. L'imperativo della comunità scientifica è pertanto l'organizzazione sistematica di questa enorme quantità di dati, per renderla utilizzabile e per generare nuove conoscenze, permettendo da un lato di giungere alla corretta interpretazione del significato patogeno delle varianti genomiche in relazione al fenotipo clinico e, all'altro di sostenere la ricerca clinica, sia essa dedicata a specifici geni o varianti genomiche oppure onnicomprensiva rispetto a geni umani e malattie. L'approccio multidisciplinare, che, a partire dalle conoscenze di base sui sistemi biologici, combina dati genomici e fenotipi, sembra essere il più utile a chiarire completamente i processi patogenetici sottostanti alle malattie e a stabilire il contributo delle singole varianti genomiche a tali processi. In tal senso sembra fondamentale la necessità di disporre di database di grandi dimensioni idonei all'acquisizione dei dati genomici ed all'analisi delle complesse relazioni con profili fenotipici d'interesse, possibilmente con identificazione di una correlazione univoca fenotipo-genotipica, e di sviluppare nuovi sistemi bio-informatici standardizzati in grado di garantire l'interoperabilità e la condivisione dell'informazione a livello globale. L'obiettivo finale è l'utilizzo della nuova conoscenza per il raggiungimento di reali benefici in termini diagnostico-assistenziali e terapeutici per i pazienti affetti da malattie rare.

Sistemi di raccolta dell'informazione genotipica, database di varianti genomiche responsabili di malattie ad ereditarietà mendeliana, giocano oggi un ruolo chiave in ambito di ricerca ed assistenza sanitaria; il concetto di mutazione genetica è noto fin dal 1949, ma la necessità di raccogliere ed organizzare il dato genotipico è nata agli inizi degli anni '60, quando fu possibile disporre per la prima volta di nuova conoscenza in ambito di genetica umana, con l'identificazione della prima mutazione genetica a livello del suo prodotto proteico (gene della globina)⁴⁵.

In questo contesto si è nel tempo creata una grande varietà di risorse di alta qualità, che integrano l'informazione genotipica, fenotipica e biologica, una serie di database, molto eterogenei per struttura, scopo, quantità e qualità d'informazioni contenute, che possono essere direttamente estratte da ricerche cliniche o di laboratorio, ricavate in modo indiretto dalla letteratura scientifica peer-reviewed o entrambi i casi. Frequentemente le banche dati sono realizzate grazie al lavoro di singoli ricercatori o di gruppi di studio interessati alla variabilità genomica umana; d'altro canto esistono anche

specifiche istituzioni o consorzi che provvedono alla creazione ed al sostentamento di banche dati specifiche per gene o malattia nel loro ambito d'interesse; ne è un esempio la banca dati InSiGHT, Società Italiana per i Tumori Gastrointestinali Ereditari⁴⁶.

Le banche dati d'informazione genotipica possono essere sistemi progettati per gestire informazioni riguardanti molteplici geni e malattie e saranno allora definibili "database universali"; tra i database universali ne esistono di due differenti tipologie: database locus-specifici e database centralizzati.

I "Locus Specific Databases" (LSDBs)⁴⁷ si focalizzano su varianti genomiche ad effetto patogeno noto e si pongono come obiettivo la copertura di tutte le varianti note per i singoli geni o loci genomici. Banche dati di questo tipo sono state istituite ad esempio per la raccolta di tutte le varianti genomiche legate ad una malattia specifica: caBIG, PMutDB, AutDB, e IDbases, banche dati per i tumori, il morbo di Parkinson, l'autismo e l'immunodeficienza, rispettivamente. Leiden Open Variation Database (LOVD), recentemente aggiornato alla terza versione, che fornisce ampi dettagli anche sul profilo fenotipico dei pazienti affetti⁴⁶. Ancora MITOMAP, MitoDat, e mtDB forniscono varianti genomiche di geni mitocondriali.

Come detto diversi database di varianti genomiche sono invece stati pensati fin dall'inizio come banche dati generalizzate di tutti i geni umani e delle varianti correlate e vengono pertanto definiti centralizzati o "General Mutation Database" (GMDB). La più grande di queste risorse è il "Single Nucleotide Polymorphism" database (dbSNP)²⁰ del National Center for Biotechnology Information (NCBI), che contiene oltre 50 milioni di varianti genetiche uniche, derivanti da dati di singole ricerche, continuamente aggiornati anche nell'ambito del Progetto Genoma 1000, uno studio che si pone come obiettivo la classificazione di oltre il 95% delle varianti nelle regioni genomiche codificabili con le attuali tecnologie di sequenziamento. Tuttavia, la corrispondenza genotipo-fenotipica nella maggioranza dei casi rimane sconosciuta: meno del 3% delle varianti esoniche missense, contenute in dbSNP, hanno un effetto patogeno noto⁴⁸.

OMIM, come detto, è un oggi un database autorevole ed esaustivo delle prime e più rimarchevoli varianti alleliche di geni nucleari e di fenotipi genetici, che raccoglie e organizza il dato genotipico disponibile nelle pubblicazioni ufficiali della letteratura scientifica peer-reviewed (Human Mutation), aggiornato quotidianamente ed

organizzato a cura della School of Medicine presso la John Hopkins University³². Nel 2013 conteneva 16.069 voci relative a varianti genomiche correlate a malattie ereditarie. OMIM garantisce un'eccellente mappatura incrociata di dati provenienti da diversi database genomici e rappresenta pertanto la risorsa più attendibile nel processo d'identificazione delle entità univoche fenotipo-genotipiche³¹. Tuttavia in OMIM sono riportate le varianti genomiche (alleleliche) ad effetto patogeno noto ed una limitata serie di polimorfismi statisticamente correlati a specifici profili fenotipi; in tal senso presenta alcuni limiti in termini di completezza e numerosità di varianti genomiche catalogate. Le collaborazioni di Orphanet hanno garantito la mappatura dei termini della nomenclatura OMIM, il cui scopo è la categorizzazione dei fenotipi clinici e dei rispettivi geni causativi, e che attualmente è largamente assimilabile alla terminologia Orphanet. Orphanet gestisce un'interfaccia di collegamento diretto con il database OMIM, che pur permettendo molto spesso l'identificazione della corretta corrispondenza fenotipo-genotipica, presenta tuttavia un limite d'inter-operabilità, essendo la classificazione di OMIM, non multigerarchica e solo parzialmente confrontabile con quella di Orphanet, all'interno di cui sono peraltro presenti numerose malattie a base genetica sconosciuta.

Universal Mutation Database (UMD) è un database relazionale a struttura unificata che fornisce un'ampia serie di dati utili all'analisi delle relazioni tra varianti genomiche e profilo fenotipico delle malattie⁴⁹.

Il database centralizzato ClinVar, costruito a partire da dati fondamentali di dbSNP ed anch'esso mantenuto presso il NCBI, oggi raccoglie ed organizza varianti genomiche specifiche, clinicamente o funzionalmente rilevanti, fornendo l'evidenza metodologica e scientifica del significato clinico o funzionale di una specifica variante e dettagli sul profilo fenotipico³⁰. La banca dati, liberamente accessibile come risorsa interattiva sul web da aprile 2013, è di ampia portata e contiene dati relativi al significato clinico o funzionale di più di 125000 varianti, comprese in tutto il genoma umano, dati raccolti ed integrati da risorse come OMIM, GeneReviews e diversi LSDBs⁵⁰.

Il consorzio "Universal Protein Resource" è un'iniziativa dell'Istituto Svizzero di Bioinformatica (SIB), dell'Istituto Europeo di Bioinformatica (EBI) in collaborazione

con “Protein Information Resource” (PIR), che fornisce alla comunità scientifica una risorsa centrale di raccolta delle sequenze proteiche funzionalmente rilevanti: la UniProt KnowledgeBase (UniProtKB), composta da un database principale, aggiornato mensilmente, e da diversi database complementari tra cui il Reference Clusters UniProt (UniRef) e l'Archive UniProt (UniParc)⁵¹. In particolare la sezione di Swiss-Prot del database principale contiene dati accessibili pubblicamente sul web, relativi a circa 40000 sequenze proteiche, derivate da un ampio spettro di organismi viventi e 66.723 varianti genomiche, di cui 23.229 a significato patogeno, e 37.446 varianti polimorfiche⁵².

Infine, non ultimo, “Human Gene Mutation Database” (HGMD) è un'altra risorsa centralizzata istituita e mantenuta presso l'Istituto di Genetica Medica a Cardiff, curata manualmente, che raccoglie in modo esaustivo un'ampia serie di varianti genomiche e di polimorfismi funzionali associati a malattie umane ereditarie⁵³; il database è diviso in una versione libera, o "Public" ed una versione proprietaria, o "Professional", rispettivamente rappresentati da 80,173 e 109,521 varianti genomiche di malattia con uno o più riferimenti nella letteratura peer-reviewed⁵⁴. Il database, disponibile per la consultazione sul web dal 1996, era in origine una risorsa pensata per la ricerca scientifica in ambito di meccanismi mutazionali nei geni umani responsabili di malattie ereditarie, ma nel tempo ha acquisito la funzione, molto più ampia di raccolta unificata e centralizzata delle varianti genomiche funzionali correlate alle malattie ereditarie umane.

Differenti risorse sono state specificamente create anche per raccolta di varianti genomiche con corrispondenza fenotipica già nota; tali risorse sono mantenute da gruppi di studio dedicati alla raccolta di informazioni pertinenti dalla letteratura scientifica peer-reviewed: un esempio è Pharmacogenomics Knowledgebase (PharmGKB) una risorsa accessibile on-line che raccoglie, organizza e diffonde informazioni sulle modificazioni del genoma umano in risposta ai farmaci (contiene 654 varianti genomiche legate alla risposta ai farmaci) e sulle possibili relazioni gene-farmaco e genotipo-fenotipiche nei percorsi di farmacodinamica e farmacocinetica³⁶.

Oltre alle risorse fin qui descritte, esistono numerosi database che possono essere utilizzabili in ambito di ricerca genomica; esempi di altre risorse create per

classificare le varianti genomiche in relazione al loro significato funzionale o alla corrispondenza fenotipo-genotipica sono Phen-Code, dbGaP, GWAS Central, HumDiv/HumVar, e NewHumVar/HumVarProf. Si tratta di banche dati derivative o cosiddette di collegamento, che integrano informazioni derivanti da uno o più database di caratteristiche fenotipiche e di varianti genomiche; la più rappresentativa ed ampia è la banca dati svedese VariBench, realizzata appositamente allo scopo di integrare un'ampia serie di dati di varianti genomiche, derivanti da differenti database e dalla letteratura peer-reviewed, e di fornire un punto di riferimento ed uno strumento per applicazione di calcoli e metodi computazionali complessi, utili alla validazione sperimentale delle conoscenze assunte in ambito di variabilità genomica⁵⁵.

I database centralizzati sembrano essere i sistemi più standard ed affidabili per la raccolta dell'informazione genotipica, perché dotati di campi d'applicazione e tipologia di dati meglio definiti; il limite maggiore di queste banche dati è tuttavia la mancanza di flessibilità metodologica, a favore di un rigoroso sistema di controllo. Tale limite si traduce in una generale mancanza d'informazioni dettagliate riguardanti specifiche malattie, con privilegio di una visione più globale e generalizzata sulla patologia umana. Le banche dati centralizzate provvedono inoltre all'integrazione automatica di dati provenienti da risorse esterne o da sistemi specifici contenenti dati inediti non standardizzati (non derivati da letteratura peer-reviewed) e quindi con una generale mancanza d'interoperabilità, che ne limita pertanto l'integrazione al più alto grado di completezza.

Per contro i vari LSDBs hanno negli ultimi anni assunto formati sempre più standardizzati ed interoperabili, grazie allo sviluppo di specifiche misure di controllo di qualità e diffusione di sistemi standard di nomenclatura⁵⁶, come si vedrà in seguito, e potrebbero rappresentare oggi preziose risorse per la conoscenza e la descrizione del profilo genotipico delle malattie. Tuttavia la mancanza di coordinamento e di collegamenti specifici tra i differenti database, la natura altamente dispersiva e la limitata copertura rispetto alle specifiche regioni genomiche studiate, li rende strumenti non completamente idonei, nella loro attuale forma, alla raccolta dell'informazione genotipica generalizzata. Inoltre i LSDBs devono anch'essi affrontare altre questioni gravose in termini di qualità metodologica, tra cui i problemi relativi alla provenienza

non standardizzata dei dati, che li rende al momento non utilizzabili anche a beneficio di database centralizzati²⁰.

In generale i database genomici derivano i dati da fonti molto eterogenee e spesso la nomenclatura di riferimento per molte varianti è stabilita via via dai singoli ricercatori, risultandone una molteplicità poco confrontabile di sigle o nomi differenti per singole regioni genomiche.

Una prospettiva risolutiva rispetto alla problematica d'interoperabilità dei sistemi di raccolta dell'informazione genotipica prevede un approccio di tipo olistico, e lo sviluppo di un sistema globale di accesso a tutti i database esistenti, di estrazione e ricatalogazione dei dati in essi contenuti. Per consentire l'interoperabilità tra database, il confronto tra le fonti ed un maggiore accesso ai dati, sembra fondamentale poter disporre di linee guida standardizzate, relative a:

- utilizzo di nomenclatura ufficiale di descrizione delle varianti genomiche
- definizione delle diverse ontologie o domini semantici a confronto;
- letteratura di riferimento
- metodologia di condivisione dei dati
- struttura dei database
- caratteristiche dei sistemi operativi per la gestione dei dati.

Attualmente vi è una generale mancanza di linee guida di riferimento ufficiale sulla maggior parte di questi aspetti, ad eccezione dei sistemi di nomenclatura e della letteratura scientifica.

Nell'aprile 1988, durante il primo meeting internazionale su tematiche riguardanti la mappatura ed il sequenziamento del genoma umano, è stata fondata la Human Genome Organization (HUGO)⁵⁷, un'organismo internazionale ancora oggi operante allo scopo di assistere la ricerca in ambito di genoma umano attraverso il coordinamento e la collaborazione tra ricercatori ed esponenti della comunità scientifica, di facilitare lo scambio di dati e bio-materiali rilevanti e, attraverso uno specifico programma di formazione, di favorire la diffusione delle tecnologie correlate, oltre che incoraggiare il dibattito pubblico e fornire informazioni su implicazioni scientifiche, etiche, sociali, giuridiche e commerciali dei progetti esistenti in materia di genoma umano.

La consapevolezza della reale mancanza di un sistema di raccolta sistematica delle mutazioni genomiche umane, ha portato, nel corso di un meeting a Montreal (Canada) nel 1994, alla fondazione della HUGO-sponsored Mutation Database Initiative (MDI), oggi nota come Human Genome Variation Society (HGVS), i cui membri hanno puntato fin da subito ad un approccio descrittivo uniforme e inequivocabile per la catalogazione delle varianti genomiche umane presenti in letteratura ed alla continua implementazione di sistemi standardizzati per l'estrazione dei dati dai differenti database e la condivisione dell'informazione genotipica a livello globale, sostenendo il miglioramento della flessibilità e dell'interoperabilità dei database stessi.

Nel 1996, sotto supervisione della HUGO, fu istituito lo HUGO Gene Nomenclature Committee (HGNC), responsabile ancor oggi della continua revisione delle linee guida complete per la nomenclatura dei geni umani, già presentate ufficialmente alla comunità scientifica per la prima volta nel 1979, in occasione dell'Edinburgh Human Genome Meeting (HGM)⁵⁷. L'HGNC ha proposto un sistema di nomenclatura inequivocabile, preciso e breve caratterizzato dall'associazione univoca di ciascun gene umano noto ad una sigla abbreviata e l'inserimento di tale sigla all'interno del HGNC database⁵⁸. Il database, consultabile al sito web <http://www.genenames.org>, contiene attualmente 33 000 sigle approvate, relative a geni codificanti pseudogeni, varianti RNA non codificanti, un numero limitato di siti fragili e quadri fenotipici⁵⁰. L'identificatore primario per ogni voce del database è il simbolo approvato dall'HGNC, un acronimo o un'abbreviazione del nome della regione genomica associata, a cui è assegnato anche un 'HGNC ID', che permette il monitoraggio dei dati anche a fronte di eventuali aggiornamenti nella nomenclatura per ingresso di nuove entità. Per ciascun simbolo è garantito anche il mantenimento di eventuali varie denominazioni "non approvate" dei singoli geni o regioni genomiche (pseudonimi), per mantenerne chiara l'identità e facilitarne la rintracciabilità all'interno delle risorse di letteratura scientifica. Il database contiene per ogni variante ulteriori informazioni quali localizzazione cromosomica, sequenza nucleotidica e collegamenti a pubblicazioni scientifiche e ad una varietà di risorse esterne. I simboli approvati sono riportati nei principali database genomici (OMIM, SWISS-PROT, HGMD, GDB, GENATLAS, GeneCards)⁵⁹.

Il sistema di nomenclatura proposto da HGVC è oggi largamente utilizzato e la HGVS ha proposto inoltre una serie di linee guida sulle caratteristiche standard minime, necessarie per un efficiente database di varianti genomiche. Tali caratteristiche comprendono, oltre all'utilizzo della nomenclatura HGNC per ciascuna variante genomica, l'associazione della stessa ad una sequenza genomica nota di riferimento, il collegamento diretto alle fonti bibliografiche specifiche per la variante e una serie d'informazioni aggiuntive riguardanti il profilo fenotipico associato.

Come detto, la realizzazione di un sistema globale di accesso e condivisione dell'informazione genotipica, ovvero la possibilità d'interrogare molteplici differenti database, utilizzando un'interfaccia ed un linguaggio comuni, è un obiettivo primario e condiviso della comunità scientifica in materia di genoma umano e resta ad oggi una sfida aperta.

Nel giugno 2006 è partito lo Human Variome Project (HVP), (HVP, <http://variomeproject.org/> //www.human), sostenuto da un'organizzazione internazionale non governativa no-profit, che punta alla catalogazione di tutta l'informazione genotipica umana, ovvero all'integrazione, conservazione, interpretazione e condivisione delle informazioni sulle varianti genomiche umane, in relazione a malattie ereditarie e risposta ai farmaci⁶⁰. HVP ha stabilito e mantiene relazioni operative con l'UNESCO e l'OMS, i suoi membri sono individui, enti, governi o organizzazioni provenienti da 81 paesi, interessati all'effettiva condivisione di informazioni sulle varianti genomiche ed in possesso di esperienza in ambito di genetica clinica e counselling genetico, di sistemi di raccolta e/o di gestione dati, d'information technology e gestione di progetti scientifici.

L'obiettivo del progetto è una collaborazione globale per raccolta dell'informazione genotipica e della sua corrispondenza fenotipica (il termine 'variome' si riferisce appunto alla corrispondenza genotipo-fenotipica nella popolazione umana) e, quindi, una conoscenza completa della variabilità genetica umana e della sua relazione potenziale con le malattie⁶¹. Gli obiettivi specifici sono lo sviluppo e l'adozione di norme standardizzate, la definizione di un consenso e di un sistema di protezione dati efficace e dotato degli adeguati requisiti etici, lo sviluppo di sistemi per il data mining automatizzato, per la partecipazione, la conservazione, la ricerca ed il recupero di dati,

l'implementazione di programmi educativi ed informativi a beneficio della comunità, anche a livello dei paesi in via di sviluppo e la promozione del concetto di medicina basata sull'evidenza scientifica.

Per raggiungere questi obiettivi, HPV collabora per la definizione di standard e linee guida internazionali, che descrivano la best-practice in ciascuna di queste aree, utilizzabile per la realizzazione di sistemi e processi specifici. Ad oggi sono state pubblicate linee guida che definiscono i parametri di qualità specifici necessari per i database gene-malattia e sono in fase di attuazione le linee guida per la valutazione dei database esistenti⁴⁷.

Strategie appropriate andranno sviluppate per il coordinamento delle attività di clinici, ricercatori, bioinformatici e curatori di database e di software adeguati alla raccolta, senza soluzione di continuità ed a basso costo, di dati provenienti da laboratori, cliniche e cartelle cliniche in tutto il mondo, alla loro catalogazione all'interno di database con caratteristiche appropriate e la loro interpretazione in relazione alla corrispondenza genotipo-fenotipica. A livello internazionale sono in atto enormi sforzi per determinare algoritmi e sistemi in grado di stabilire a priori le probabilità che una variante genomica si associ ad un dato effetto patogeno, in base alle associazioni genotipo-fenotipiche note ed a caratteristiche chiave, quali i processi evolutivi di conservazione del DNA, la natura e la frequenza delle varianti missenso e dei cambiamenti strutturali dei prodotti proteici. Un unico database internazionale potrebbe fornire informazioni complete e accurate su varianti genomiche e polimorfismi a singolo nucleotide, correlate a dati fenotipi derivanti da studi di associazione. La raccolta effettiva di dati sperimentali e il loro posizionamento nel database HVP potrà avvenire attraverso un'infrastruttura centralizzata con un modello comune, un'interfaccia unica e con specifiche funzionalità di ricerca. Il mantenimento di questa infrastruttura è garantito da collaborazioni a livello internazionale, mentre i singoli LSDBs, con i massimi esperti di una data malattia e curatori dedicati, rappresentano la risorsa necessaria alla cura di sezioni locus-specifiche, contenenti informazioni fino al massimo dettaglio sulla variante o regione genomica in questione. Vi è anche la necessità di un accordo sulla governance dei LSDBs a livello internazionale, per un adeguato finanziamento e per la garanzia che l'umanità raccolga i benefici della corrente rivoluzione genetica molecolare.

Un approccio coordinato a livello internazionale e sostenuto da organizzazioni ed esperti autorevoli per malattia sembra auspicabile, ma impegna una vasta gamma di professionisti e risorse eterogenee. Attualmente sono in via di sviluppo strumenti e sistemi bio-informatici in grado di integrare differenti ontologie o domini di conoscenza e linee guida che regolamentano la fornitura d'informazioni sulle varianti genomiche ed il passaggio delle stesse dai LSDB alle banche dati centralizzate.

In conclusione HVP sembra poter garantire la realizzazione di questi obiettivi; ad oggi le linee guida e le normative stabilite da HVP rappresentano il gold standard per la comunità scientifica. In alcuni ambiti devono essere ancora identificate le linee guida di riferimento, mentre sono in fase di sviluppo sistemi bio-informatici per la raccolta e la gestione dell'informazione genotipica a livello globale, che dovrebbe partire dalle risorse disponibili nei LSBs e da gruppi di lavoro specificamente dedicati a singole patologie o gruppi di patologie e raggiungere in modo standardizzato le banche dati centralizzate, con buone prospettive future⁶².

Progressi recenti hanno indicato che la raccolta di varianti genomiche a livello mondiale è teoricamente possibile. Un altro esempio è il progetto pilota per la raccolta di tutte le varianti esistenti delle quattro regioni genomiche mismatch-repair, che sono alterate nei pazienti affetti da cancro del colon, avviato dalla Società Internazionale per i Tumori Gastrointestinali Ereditari (Insight).

1.4 SISTEMI DI MONITORAGGIO SANITARIO

1.4.1 Registri per le malattie rare

Da quanto emerso sul profilo epidemiologico, sulle attuali criticità dei sistemi classificativi e di raccolta dell'informazione fenotipica e genotipica in ambito di malattie rare, sembra fondamentale poter disporre nel prossimo futuro di una reale e solida conoscenza sull'impatto di queste malattie a livello di popolazione, sul loro decorso e sulla prognosi a lungo termine. Tale conoscenza permetterà di adeguare le politiche sanitarie ai bisogni dei malati rari, di guidare le decisioni politiche ed i finanziamenti e quindi di attuare, in ambito di salute pubblica, strategie idonee a garantire ai pazienti affetti l'accesso ad un percorso assistenziale diagnostico-terapeutico di elevata qualità. Attualmente l'assenza di un sistema di classificazione e codifica delle malattie rare, riconosciuto a livello internazionale, rappresenta ancora un

vero limite alla ricerca clinica, che necessita di progetti mirati e, per la corretta valutazione del reale impatto economico sociale e di salute pubblica delle malattie rare, deve avvalersi di specifiche e costose risorse dedicate alla raccolta di dati epidemiologici per le singole patologie o gruppi di patologie o per limitati bacini di popolazione o aree geografiche.

A livello di Comunità Europea (CE), la specificità delle malattie rare le configura oggi come un ambito di grande interesse, nel quale la cooperazione tra gli stati rappresenta un alto valore aggiunto per massimizzare le scarse risorse fin qui destinate e ottenere misure epidemiologiche che stimino l'impatto globale delle malattie rare in Europa ed adeguare i piani sanitari⁶³. L'attenzione dei ricercatori, dei pazienti e delle istituzioni governative si sta rivolgendo alle strategie più idonee alla raccolta sistematica d'indicatori epidemiologici in ambito di malattie rare⁶⁴.

In base ai dati emergenti dalla letteratura, i registri di pazienti e di patologia costituiscono strumenti interessanti per lo sviluppo delle conoscenze epidemiologiche e della ricerca clinica; altresì l'individuazione di centri di riferimento esperti in ambito di malattie rare e la creazione di reti di collaborazione a livello nazionale ed internazionale sono necessarie per ampliare le conoscenze sul profilo epidemiologico, per l'adeguamento delle cure e per il miglioramento della programmazione sanitaria in ambito di malattie rare⁶⁵.

I primi dati derivati da registri di malattia risalgono ai primi anni '90, con la pubblicazione del registro nazionale statunitense per l'ipertensione arteriosa polmonare, anche se già a partire dagli anni '70 erano attivi alcuni registri; nei primi anni 2000 la comunità scientifica inizia a riconoscere e discutere la necessità di sviluppare registri clinici, multicentrici in ambito di malattie rare⁶⁶.

Registri europei ed extra-europei

Secondo il più recente report di Orphanet di gennaio 2015, in Europa sono attualmente attivi 651 registri, con copertura rispetto a più di 800 patologie rare, nell'82% dei casi sostenuti da risorse pubbliche e in circa la metà dei casi stabiliti per iniziativa di medici, ricercatori o associazioni di pazienti⁶⁷.

La maggior parte (60,5%) dei registri attualmente esistenti in Europa sono stati istituiti tra il 2000 ed il 2011, quando la comunità scientifica ha iniziato a riconoscere e

discutere la necessità di sviluppare registri clinici, multicentrici in ambito di malattie rare.

Registri di malattia rappresentano il 34,1% del totale, mentre registri dedicati a gruppi di patologie tra loro correlate rendono conto del 46,6%. Registri che hanno una ampia copertura rispetto alla totalità delle malattie rare sono circa il 7,3%, la maggior parte dei quali (56,3%) contenente più di 5000 pazienti registrati, anche se non esiste correlazione tra la copertura geografica ed il numero di pazienti⁶⁸.

La letteratura scientifica individua alcuni principali, ricorrenti obiettivi per i quali gli esistenti registri a livello europeo sono stati istituiti:

- nel 70,8% dei casi si tratta di registri per ricerca a scopo epidemiologico;
- il 61,2% dei registri rappresentano una risorsa di reclutamento pazienti per trial clinici;
- nel 60,7% dei casi sono stati creati per aumentare le conoscenze rispetto alla storia naturale di malattia o per sorveglianza clinica rispetto ad una patologia (55,5%);
- il 53,4% risulta istituito a scopo di stabilire una correlazione genotipo-fenotipo;
- nei restanti casi per differenti ragioni, quali pianificazione dei servizi sanitari o sociali, valutazione dell'efficacia e monitoraggio sulla sicurezza dei trattamenti etc.

Rispetto alla copertura di popolazione, in europa il 61,9% dei registri si colloca a livello nazionale, mentre solo il 18% ha una copertura europea e internazionale. Più della metà sono registri di popolazione (56,7%), soprattutto nel caso di registri a copertura nazionale, mentre i registri locali o regionali contengono per lo più dati derivanti dalle strutture ospedaliere. I registri a copertura europea ed internazionale riportano prevalentemente dati derivanti da studi familiari o di coorte.

I dati contenuti nell'83,6% dei registri derivano da risorse in ambito clinico, mentre il 43,8% circa da risorse in ambito genomico e sono nel 95% dei casi dati relativi alla diagnosi di malattia, nell'86,8% dati clinici e nel 72,3% dati genomici.

Molti registri esistenti (36,5%) non utilizzano codici per l'identificazione di malattia, negli altri casi emerge l'utilizzo di codici ad hoc, oppure il ricorso a differenti terminologie di riferimento internazionale, quali l'ICD (versioni 9 e 10), le nomenclature ORPHA, MIM, SNOMED-CT and MESH. La quasi totalità dei registri (95%) viene costantemente aggiornata e riporta periodici controlli di qualità sui dati contenuti secondo molteplici procedure, ma circa un quarto non riporta la data di morte del paziente. Le principali risorse economiche risultano stanziare a favore dei registri per i cui risultano procedure stabilite di controllo qualità, affidabilità, accordo e validità strutturale; alcuni registri, che consentono l'accesso ai propri dati adottano come più frequente procedura di controllo per l'accesso ai dati stessi l'approvazione da parte di un comitato interno (52%) o, in alternativa, specifici accordi contrattuali. La sicurezza dei dati è garantita comunemente attraverso l'utilizzo di un server dedicato, protetto (78,2%) o con utilizzo di strategie di backup dei dati (62,9%) e di password temporanee (58,9%).

Circa il 25% dei registri europei sono stati creati in assenza di specifiche dedicate risorse economiche e, superata la fase iniziale, molti registri operano attualmente senza stanziamenti finanziari, risultando difficilmente sostenibili da parte delle associazioni di pazienti, delle università, degli istituti di ricerca e della CE stessa. Solo il 51,3% dei registri europei dispone ad oggi in ambito di politiche sanitarie di un piano strategico di sostenibilità a lungo termine. Circa il 50% dei registri europei ha attualmente aperte collaborazioni con altri registri cui condivide i propri dati. La maggioranza dei registri europei (73%) si dichiara a favore di una piattaforma unica che coordini tutti i registri esistenti, ovvero di un'infrastruttura globale che consenta di disporre di servizi e strumenti comuni⁶⁸.

Anche in ambito extra-europeo i registri sono stati identificati come strumenti particolarmente utili per la rintracciabilità e la caratterizzazione del profilo epidemiologico delle malattie rare, pur con le medesime criticità emerse a proposito dei registri europei⁶⁹.

Per i malati rari negli Stati Uniti sono attualmente esistenti numerosi registri di pazienti e di malattia, anch'essi molto eterogenei per struttura, scopo e caratteristiche generali. Le necessità percepite a livello di comunità scientifica riguardano soprattutto

la possibilità di disporre di misure epidemiologiche a livello di popolazione generale ed estendere la conoscenza sull'impatto delle malattie rare a livello globale.

Nel 2010, in occasione del un workshop “Advancing Rare Disease Research: the Intersection of Patient Registry, Biospecimen Repositories and Clinical Data” organizzato a Bethesda dal “Office of Rare Diseases Research (ORDR)”, dal “National Eye Institute” e dal “National Center for Research Resources dell'NIH si è istituito un movimento, che ha proposto e discusso la creazione di un registro globale dei pazienti per le circa 7.000 malattie rare conosciute “The Global Rare Diseases Patient Registry and Data Repository (GRDR)”. L'istituzione coinvolta in prima linea è l'ORDR, un organismo istituito presso l'NIH nel 2002 allo scopo di finanziare, promuovere e coordinare la ricerca in ambito di malattie rare, con creazione di standard operativi utili alla raccolta sistematica ed alla condivisione dell'informazione⁷⁰.

Il workshop ha visto la partecipazione di diverse centinaia tra ricercatori, clinici e bio-informatici, pazienti ed associazioni, enti governativi ed istituzioni pubbliche, industrie ed aziende farmaceutiche, la cui attenzione si è rivolta alla creazione di un'infrastruttura basata sul Web, predisposta all'archiviazione d'informazioni cliniche sui pazienti affetti da malattie rare negli stati uniti, anche attraverso la raccolta incrociata di dati provenienti dai vari registri esistenti. A tal proposito è stata anche considerata la possibilità di collegare al registro globale una banca dati centralizzata di campioni biologici di malati rari, ovvero una Biospecimen Repository, (RD-HUB) <http://biospecimens.ordr.info.nih.gov/>. Il registro si pone come obiettivo l'implementazione del processo di raccolta dati aggregati provenienti da fonti molteplici attraverso la creazione di un standard comune, per consentire la ricerca traslazionale in ambito di malattie rare aumentando la conoscenza sulla storia naturale delle stesse e facilitando la rintracciabilità ed il reclutamento dei pazienti in studi clinici. Le tematiche discusse hanno riguardato norme e regolamenti governativi in materia di privacy e diritti dei pazienti, la natura, la portata e l'accuratezza dei dati potenzialmente contenibili nel registro globale, modelli e sistemi internazionali di reclutamento dei pazienti, nonché metodologie di aggregazione dei dati provenienti da fonti eterogenee, sistemi di codifica, di accesso e protezione dei dati stessi.

Lo sviluppo del GRDR rappresenta l'espressione massima dell'impegno dell'intera comunità scientifica a favore delle malattie rare e della volontà di garantire ai pazienti affetti, come mai è avvenuto in passato, l'adeguatezza del percorso diagnostico e di cura e il miglioramento della qualità di vita.

Il registro sarà disponibile a completamento di uno studio pilota in corso in questi anni.

I dati attualmente disponibili in letteratura scientifica sulla restante situazione dei registri per le malattie rare a livello internazionale restano ancora limitati e frammentari.

Da annoverare il registro globale canadese per l'“Atypical hemolytic uremic syndrome (aHUS), progettato ed istituito nel 2012 per la raccolta prospettica, multicentrica d'informazioni demografiche ed epidemiologiche riguardanti tutti i pazienti eleggibili affetti da aHUS⁷¹. L'esperienza del registro canadese ha sottolineato l'importanza dei registri di malattia per la raccolta d'indicatori utili, quali la storia naturale di malattia, le correlazioni fenotipo-genotipiche, l'esito dei trattamenti, la sicurezza delle terapie farmacologiche e la qualità delle cure. I dati contenuti si sono mostrati potenzialmente utili anche nella definizione di strategie future per ottimizzare il percorso assistenziale terapeutico del paziente affetto, favorendo la partnership tra la comunità accademica e di ricerca e l'industria globale.

In Australia esistono numerosi registri per le malattie rare, ma quelli dotati di elevati standard di qualità e buona copertura di popolazione sono limitati: solo cinque registri hanno una copertura a livello nazionale⁷². Uno di questi, l'“Australia and New Zealand Dialysis and Transplant Registry (ANZDATA)”, è una risorsa preziosa a livello nazionale, perché raccoglie i dati su tutti i pazienti che hanno ricevuto un trapianto di rene o sono sottoposti a dialisi renale e fornisce indicatori utili sul rischio di salute, sulla conformità delle cure alle linee guida di riferimento, sui problemi di accesso alle cure stesse e sull'equità nell'erogazione dei servizi, consentendo la pianificazione delle risorse ed il controllo sulla pratica clinica. L'obiettivo, anche per l'Australia è l'implemento dei registri esistenti per renderli di portata nazionale e la creazione di nuovi registri per disporre di misure di qualità valide e affidabili, confrontabili tra diverse istituzioni a livello nazionale ed internazionale.

L'esperienza asiatica in ambito di registri per malattie rare è scarsa e costellata di criticità⁷³; in Cina il primo registro per malattia rara è stato istituito solo nel 2010 dalla “China-Dolls Care and Support Association”, un'associazione di pazienti affetti prevalentemente da osteogenesi imperfetta. Il registro è basato su una piattaforma web, predisposta per la raccolta d'informazioni cliniche ed accessibile a medici e pazienti; il modello ha tuttavia evidenziato notevoli criticità strutturali, di operabilità e di funzionamento, e caratteristiche non compatibili con gli standard identificati a livello europeo e statunitense per i registri per le malattie rare⁷⁴.

Uno studio del 2012 riporta misure di prevalenza delle malattie rare in Taiwan, provenienti dal registro globale per le malattie rare, istituito nel 2002, sottolineando tuttavia l'inconsistenza degli indicatori epidemiologici riportati, notevolmente sottostimati rispetto alle misure di prevalenza negli altri paesi, e quindi la scarsa qualità dei dati raccolti dal registro, con necessità di un maggior coinvolgimento degli organismi governativi e delle autorità sanitarie nel monitoraggio delle malattie rare a livello di popolazione⁷.

Registro Nazionale per le Malattie Rare (RNMR)

In Italia sono ad oggi stati attuati diversi Piani Sanitari Nazionali (PSN), 1998-2000, 2003-2005 e 2006-2008, che hanno da subito indicato come priorità la tutela dei malati rari; in particolare, il Piano Sanitario Nazionale 1998-2000 individuava tra gli interventi prioritari “l'identificazione di Centri Nazionali di riferimento per patologie e la costituzione di una rete di Presidi Ospedalieri, a essi collegati, per la diagnosi e il trattamento di patologie rare (singole o gruppi)”.

Nel 2001 è stato emanato il Decreto Ministeriale 279/2001, che, in ambito di misure dedicate al miglioramento del processo clinico-assistenziale dei malati, rari prevedeva:

- la realizzazione di una rete clinico-epidemiologica, costituita da Presidi Accreditati, individuati dalle regioni e formalmente designati perché riconosciuti come competenti per la prevenzione, la sorveglianza, la diagnosi e il trattamento delle malattie rare e per l'esclusiva disponibilità di servizi complementari;

- l'istituzione presso l'Istituto Superiore di Sanità (ISS) del Centro Nazionale per le Malattie Rare (CNMR) e di un Registro Nazionale per le malattie rare, (RNMR).

Il mandato è stato quello di monitorare la situazione globale delle patologie rare in Italia, favorendo il flusso di informazioni clinico-epidemiologiche, dai centri regionali di riferimento, e quindi dai Registri Regionali (RR) al RNMR e sostenendo così la pianificazione delle politiche sanitarie. Il decreto stabiliva altresì la liceità della collaborazione con altri registri internazionali, per facilitare la raccolta di ulteriori dati demografici, anamnestici, clinici e di laboratorio, necessari a scopo epidemiologico ⁷⁵.

La rete si è strutturata aderendo al sistema organizzativo del sistema sanitario nazionale italiano, che decentra la responsabilità di erogazione e gestione dei servizi sanitari alle Regioni, ed il RNMR risulta funzionalmente integrato ai RR, stabiliti dalle autorità regionali. Nel 2002 e nel 2007 sono stati realizzati due accordi Stato-Regioni, che hanno assicurato l'attuazione coordinata di tutte le unità della rete clinico-epidemiologica dei RR e del RNMR, definendo variabili e procedure per la libera circolazione, la standardizzazione e la condivisione delle informazioni cliniche e dei dati. In particolare sono state individuate le normative necessarie per garantire la legittimità e la totale sicurezza del flusso di dati, in accordo alle legge sulla protezione dei dati personali. Il passaggio dei dati dai RR al RNMR è regolamentato da un algoritmo specifico per la protezione dei dati, che sono trasferiti attraverso appositi canali transitoriamente ed esclusivamente aperti in rete; dati personali ed dati sensibili sono gestiti su server distinti dell'ISS, protetti da avanzati firewall e da sistemi tecnologici.

Le autorità sanitarie nazionali e regionali hanno costantemente collaborato per la realizzazione e lo sviluppo di questa complessa struttura in rete di Presidi Accreditati a livello regionale e per la progressiva istituzione dei RR. Tra il 2002 ed il 2011 tutte le regioni, in tempi diversi, hanno istituito un RR; alcune hanno predisposto da subito sistemi informativi specifici per la raccolta e l'archiviazione dei dati ed in alcuni casi anche per la gestione assistenziale del malato raro. Fino al 2009 sono stati condotti da parte del CNMR corsi di formazione alle Regioni, procedure di standardizzazione dei dati e di controllo di qualità.

Il sistema ha raggiunto la totale copertura del territorio nazionale nel corso del 2011 e nel giugno 2012 il RNMR contava 110,841 record di malati rari; le procedure di raccolta dati raccolti sono state costantemente sottoposte a misure di controllo di completezza e qualità e nel corso degli anni 2000 si è assistito ad un progressivo miglioramento qualitativo dei dati raccolti, in termini di completezza rispetto alla diagnosi, alle manifestazioni cliniche, all'esordio ed al decorso di malattia, nonché dei farmaci orfani implicati.

Le attività di coordinamento del RNMR con altri registri in ambito europeo sono state gestite dall'ISS attraverso il CNMR (<http://www.cnmr.iss.it>), anche con collaborazioni e progetti con servizi nazionali di statistica, medici, associazioni di pazienti, tramite la piattaforma europea per i registri delle malattie rare (EPIRARE: www.epirare.eu)¹², l'International Rare Disease Research Consortium, (IRDiRC: www.irdirc.org) e la piattaforma integrata di collegamento per database, registri, biobanche e sistemi bioinformatici per la ricerca in ambito di malattie rare (RD-Connect: www.rd-connect.eu).

Registro per le Malattie Rare del Veneto (RMRV)

Il Registro per le Malattie Rare del Veneto è un sistema informatizzato di monitoraggio delle malattie rare attivo dal 2002 nella Regione Veneto, che presiede alla registrazione dei malati rari ed alla gestione del paziente all'interno di una rete clinico-assistenziale, attraverso il collegamento dei Centri Accreditati (CA) per le malattie rare, delle Aziende ULSS, dei Distretti Socio-Sanitari e dei Servizi Farmaceutici.

Tra gli obiettivi del registro si annoverano la generazione di informazioni epidemiologiche essenziali per orientare le politiche ed i piani di programmazione sanitaria ed i finanziamenti pubblici, descrivere la storia naturale ed assistenziale del malato raro, attraverso la stratificazione di un'informazione clinica completa e sempre disponibile, e la creazione di una banca informativa comune a tutta la rete dei CA, per lo sviluppo in ambito di ricerca clinica avanzata.

In dieci anni d'attività il RMRV ha consentito di disporre di utili misure epidemiologiche, derivanti da un ampio bacino di popolazione e su un eterogeneo gruppo di malattie rare non correlate¹⁴.

Vantaggi e criticità dei registri per le malattie rare

Come visto vi è ormai un certo accordo, a livello di comunità scientifica internazionale, nel riconoscere i registri per le malattie rare, soprattutto i registri di pazienti a copertura nazionale, come strumenti vantaggiosi per diversi aspetti sia in ambito di ricerca clinica, sia rispetto alla dimensione assistenziale terapeutica.

I registri permettono la migliore diffusione della conoscenza sulle malattie rare perché rendono disponibili informazioni a ricercatori e pazienti, utili a stabilire nuovi protocolli di sperimentazione che ottimizzino gli standard terapeutici, migliorino la pratica clinica e aiutino la comunità scientifica ad orientarsi su problemi inerenti la sicurezza dei pazienti ed il reclutamento degli stessi in studi clinici. In ambito epidemiologico clinico forniscono inoltre attendibili misure statistiche descrittive a livello di popolazione, quali indicatori di prevalenza ed incidenza e sulla storia naturale delle malattie.

In ambito strettamente clinico rappresentano uno strumento efficace per la reale valutazione degli esiti d'interventi preventivi, diagnostici, terapeutici e per il monitoraggio della sicurezza dei farmaci orfani, che non sono generalmente testati in ampi studi di fase 3 e che quindi necessitano di una sorveglianza post-marketing ancora più imponente dei farmaci convenzionali. I registri forniscono inoltre valide misure globali di outcome, attraverso la standardizzazione e la riduzione della variabilità dei dati derivanti dalla pratica clinica.

Attraverso la corretta raccolta dell'informazione clinica fenotipo-genotipica e l'analisi dei dati disponibili nei registri sembrano possibili studi di associazione fenotipo-genotipo e studi di linkage con campioni e tessuti biologici per la rilevazione di correlati fenotipici.

In particolare in ambito europeo il Consiglio della CE, come visto, nelle raccomandazioni pubblicate su "Official Journal of The European Union" l'8 giugno 2009, ha riconosciuto come priorità la codifica e la rintracciabilità delle malattie rare nei sistemi informativi d'assistenza sanitaria, che sembrano necessarie per disporre di adeguati indicatori epidemiologici a livello di popolazione. Le indicazioni riguardano la necessità di migliorare il coordinamento dei programmi comunitari, nazionali e regionali per la ricerca in ambito di malattie rare, sostenere l'importanza di specifiche risorse, quali registri di malattia, a fini epidemiologici, identificare centri di alta

specializzazione e competenza specifica e promuovere la creazione di Reti Europee di Riferimento per malattie rare. In particolare, EUCERD, tra il 2004 e il 2013 è stato incaricato di assistere la CE nel definire le politiche sanitarie in materia di malattie rare e d'incoraggiare gli Stati membri ad elaborare piani e strategie nazionali con interventi pertinenti nel settore sanitario e sociale¹².

EUCERD ha assistito la CE nell'elaborazione di raccomandazioni relativamente a specifici ambiti:

- i criteri di qualità per la selezione di Centri di Eccellenza esperti di malattia;
- le decisioni in ambito di farmaci orfani, sostenute dall'evidenza clinica;
- la struttura ed organizzazione di Reti Europee di Riferimento (RER) per le malattie rare;
- la raccolta di indicatori epidemiologici e la rintracciabilità dei malati rari nei sistemi sanitari informativi.

Nel report del 2011 EUCERD⁶³ ha ribadito anche le criticità e le aree d'intervento per fare fronte alla necessità d'indicatori clinico-epidemiologici in ambito di malattie rare, soprattutto per valutare lo stato di salute e la prognosi dei malati rari e l'efficacia delle strategie operative e delle politiche sanitarie ad essi rivolte a livello nazionale ed europeo. Il gruppo di lavoro si è dedicato allo specifico ambito dei registri di malattia rara come potenziali fonti d'indicatori di qualità per la cura delle malattie rare, attraverso la conoscenza approfondita del profilo epidemiologico delle stesse e la conseguente pianificazione sanitaria. La priorità è la validazione dei registri di malattia secondo criteri di qualità standardizzati, che riguardino caratteristiche quali la dimensione del campione di popolazione e la copertura geografica, le metodologie d'acquisizione e gestione dati e l'interoperabilità tra differenti registri, necessaria per la condivisione ed il confronto dei dati.

Nel 2011 la CE ha istituito EPIRARE (“Building Consensus and Synergies for the EU Registration of Rare Disease Patients”, www.epirare.eu), ⁷⁶un progetto pensato con l'obiettivo di stabilire i requisiti necessari per registri in ambito di malattie rare ed indagare un modello di piattaforma globale per i registri europei, che fornisca servizi e strumenti comuni per la condivisione di dati accuratamente standardizzati. Nel progetto

sono stati definiti una serie di elementi comuni per la raccolta d'informazioni sulla qualità dei registri ed altri indicatori d'interoperabilità tra gli stessi e di tracciabilità dei dati e grazie al progetto è stato possibile disporre di una visione globale sulla reale situazione dei registri attualmente presenti in europa. Nel report EUCERD del 2012 ⁷⁷; i registri sono stati definitivamente riconosciuti come unica risorsa utilizzabile per disporre di dati sufficienti alla ricerca epidemiologica e/o clinica, come importanti per valutare la fattibilità degli studi clinici, per facilitare la pianificazione sanitaria e per sostenere l'arruolamento dei pazienti.

I registri di malattia attualmente presenti in Europa sono, nella stragrande maggioranza dei casi, basati su studi di popolazione e dedicati ad una singola patologia o ad un gruppo di patologie correlate, come risultato di iniziative spontanee, quali ad esempio singoli progetti di ricerca. I registri sono prevalentemente sostenuti da un consiglio direttivo ed i dati in essi contenuti derivano quasi esclusivamente da centri clinici o strutture assistenziali. Questi aspetti riflettono l'interesse generale della comunità e la volontà di disporre di maggiori informazioni sulla salute dei malati rari; un approccio promettente per ottenere validi indicatori epidemiologici, oltre che informazioni strettamente connesse all'ambito clinico, sembra essere quello basato su studi di popolazione. Molti registri devono ancora affrontare questioni legate alla qualità ed alla standardizzazione dei dati o all'ottimizzazione delle risorse impiegate nella raccolta dati, per evitare sforzi inutili o la perdita consistente d'informazioni utili. Molti registri ancor oggi non rendono accessibili o condivisibili i propri dati, non essendo ben strutturate le procedure di controllo di qualità, se non in caso di registri che ricevono finanziamenti dedicati, non esistendo disposizioni chiare in merito a questi aspetti ed essendo le norme che regolamentano le questioni etiche e legali correlate percepite come un ostacolo alla libera circolazione ed alla protezione dei dati.

Altro aspetto svantaggioso è l'assoluta frammentazione dei sistemi di codifica delle malattie, che influisce in maniera rilevante sulle caratteristiche d'interoperabilità tra registri e sul processo di condivisione delle informazioni. In tal senso le criticità di cui si è visto, in ambito di sistemi classificativi per le malattie rare, giocano un ruolo fondamentale.

Le principali necessità emergenti in ambito di registri di malattie rare riguardano pertanto le misure di controllo qualità dei dati, il miglioramento dell'accessibilità ai dati e della visibilità dei registri stessi a livello globale, oltre che lo stanziamento di risorse economiche specificamente dedicate. Il soddisfacimento delle esigenze dichiarate dai sostenitori dei registri di malattia richiedono appropriate strategie operative a livello di autorità governative e specifiche norme che regolamentino la proprietà e la partecipazione ai registri stessi; purtroppo la messa in atto di questi interventi non sembra essere a portata di mano al momento attuale. La maggioranza dei registri si dichiara a favore della realizzazione di un modello unico di piattaforma condivisa⁷⁸, che fornisca strutture e servizi comuni ai registri esistenti, per gli aspetti tecnologici e per altri aspetti strutturali necessari, non direttamente connessi all'ambito clinico, o allo scopo per cui i singoli registri sono costituiti, ma che richiedono competenze specifiche e risorse dedicate al sostegno di operatività ed inter-operabilità ai massimi livelli.

La situazione attuale dei registri appare comunque matura per la transizione verso un approccio più uniforme, coordinato e di alta qualità, che sembra portare alla costituzione di registri globali o di piattaforme comuni di coordinamento di tutti i registri esistenti e dei registri futuri. Tali infrastrutture dovrebbero svolgere funzioni di repository dei dati, di supporto alla ricerca clinica e di sostegno alla produzione di output predefiniti per regolamentare i finanziamenti pubblici e privati, la fornitura di strumenti e risorse per l'utilizzo registri, la registrazione sistematica dei pazienti e la creazione di reti tra i pazienti e professionisti.

La realizzazione di questo modello è attualmente un obiettivo strategico della comunità europea nell'ambito del progetto EPIRARE e della comunità statunitense nell'ambito del progetto sul GRDR.

1.4.2 Fascicolo Sanitario Elettronico individuale

Nel percorso evolutivo che sostiene l'implementazione del servizio sanitario a miglior tutela dell'esercizio del diritto alla salute al cittadino, sono fondamentali azioni quali l'alleggerimento burocratico, la personalizzazione delle cure e la riduzione dell'errore umano.

Nell'epoca recente si è assistito allo sviluppo di tecnologie informatiche innovative che rendono potenzialmente disponibile una grande quantità d'informazioni complete ed idonee a garantire al cittadino la migliore continuità assistenziale.

Il Fascicolo Sanitario Elettronico ("FSE") inteso come la raccolta corretta ed il più esauriente possibile dei dati clinici di un individuo si pone come obiettivo principale il miglioramento della qualità assistenziale, nel rispetto delle normative per la protezione dei dati personali.

L'11 novembre 2010 il Ministero della Salute ha emanato le Linee Guida Nazionali per il FSE, approvate in seguito in occasione della successiva Conferenza Stato-Regioni, che rappresentano il riferimento unitario nazionale per la realizzazione del FSE e in cui sono definiti, per la prima volta in atti ufficiali, i contenuti del FSE, e gli elementi necessari per una coerente progettazione ed impiego di tale struttura nell'ambito del Servizio Sanitario Nazionale (SSN) e del più ampio contesto europeo⁷⁹. Con l'introduzione del FSE è auspicabile che nel prossimo futuro, si assista ad un rivoluzionario cambiamento dei rapporti tra infrastruttura sanitaria e cittadino, fin qui caratterizzati da barriere comunicative, da procedure burocratiche macchinose e rigide e dalla documentazione cartacea .

Il contesto in cui si sviluppa il progetto del FSE riguarda i Livelli Essenziali di Assistenza (LEA), previsti dal decreto legislativo del 30 dicembre 1992, n.502, che rappresentano lo strumento del SSN per garantire al cittadino condizioni di uniformità sul territorio nazionale e sono stati istituiti con l'intenzione di creare un riferimento omogeneo a livello nazionale per i servizi sanitari, in termini quantitativi e qualitativi e nel rispetto di predeterminate risorse.

Gli obiettivi previsti nei LEA sono regolamentati secondo la vigente ripartizione delle competenze tra enti centralizzati e risorse locali, secondo i 3 livelli del SSN dotati di autonomia politico-istituzionale: Stato, Regioni ed Aziende Locali. Tra Stato e Regioni sono previsti accordi specifici per la definizione dei LEA, l'attivazione di sistemi di monitoraggio, l'individuazione delle risorse, attività e prestazioni necessarie alla tutela della salute e la definizione di linee guida per l'erogazione uniforme dei servizi. Il mantenimento dei LEA richiede pertanto anche la programmazione a livello regionale e a tal proposito, nell'accordo Stato-Regioni dell'8 agosto 2001, si è definito

l'impegno delle Regioni stesse a sostenere la spesa sanitaria necessaria a consentire l'erogazione delle prestazioni previste dai LEA.

Il FSE si definisce come “l'insieme dei dati e documenti digitali di tipo sanitario e socio-sanitario generati da eventi clinici presenti e trascorsi, riguardanti l'assistito”, è ‘luogo’ istituzionale della Rete (My Page di Internet) di archiviazione personalizzata e costantemente aggiornata nell'ambito del SSN e dei servizi socio-sanitari regionali, della storia clinica di un individuo, con un orizzonte temporale che copre l'intero ciclo di vita.

Il FSE si prevede sia costituito, previo consenso dell'assistito, per diverse finalità che vanno dall'implementazione del percorso di prevenzione, diagnosi, cura e riabilitazione ad applicazioni per la ricerca scientifica in ambito medico, biomedico ed epidemiologico ed al controllo di qualità delle cure e dell'assistenza sanitaria. I contenuti del FSE riguardano:

- dati identificativi dell'anagrafica dell'assistito;
- dati amministrativi di tipo assistenziale, costituiti da informazioni sulla posizione del cittadino nei confronti del SSN;
- documenti sanitari e socio-sanitari “certificati”, rilasciati dal SSN e archiviati elettronicamente in banche dati dedicate. Il FSE è costituito in questa sezione da un nucleo minimo di documenti indispensabili, referti, verbali Pronto Soccorso, lettere di dimissione, il Profilo Sanitario Sintetico e da una serie di documenti integrativi. Il Profilo Sanitario Sintetico costituisce il documento informatico sanitario riassuntivo della storia clinica di un individuo, creato con l'obiettivo di velocizzare il percorso di cura attraverso un rapido inquadramento del paziente anche in situazioni di emergenza e contenente un set predefinito di dati clinici significativi aggiornati.

Nell'ambito del FSE il “Taccuino personale del cittadino” è invece una sezione riservata al cittadino per inserimento “non certificato” di informazioni personali, di documenti sanitari altri (es. referti di esami effettuati in strutture non convenzionate con il SSN), diario e promemoria di eventi medici rilevanti o parametri di monitoraggio.

L'infrastruttura tecnologica globale del FSE, di cui si occupa in Italia l'Agenzia per l'Italia Digitale (AgID), provvede all'organizzazione dei sistemi di raccolta dei documenti e dei dati sanitari digitali, ovvero all'accesso ai dati da parte degli operatori del SSN, ed alla gestione dell'evoluzione temporale dei processi sanitari e dei servizi di supporto correlati; tale infrastruttura si pone inoltre l'obiettivo di compatibilità con i modelli architettuali delle infrastrutture regionali già esistenti, verso un modello d'infrastruttura federata, gestita a livello nazionale ed allineata allo scenario di riferimento internazionale. A tale proposito l'adozione di standard riconosciuti a livello internazionale risulta indispensabile ed è ormai l'orientamento principale nella maggior parte dei Paesi, nella realizzazione di corretti sistemi di codifica e classificazione dell'informazione clinica, determinanti ai fini della condivisione e interpretazione dei contenuti informativi. Tali standard devono prevedere, per tutti i sistemi classificatori, specifici criteri di qualità (esaustività, economicità, orientamento specifico, mutua esclusività etc.) e caratteristiche d'interoperabilità funzionale e semantica in diversi contesti operativi, attualmente oggetto di progetti nazionali ed europei.

L'architettura del FSE deve poter consentire la localizzazione e la disponibilità delle informazioni sanitarie, supportare adeguatamente i processi sanitari e il modello d'organizzazione del SSN, consentire una facile integrazione e interoperabilità con sistemi e infrastrutture preesistenti e garantire un elevato livello di sicurezza e affidabilità. Il FSE deve inoltre potersi integrare con il Sistema Pubblico di Connettività (SPC) ed essere conforme alle indicazioni del Garante della Privacy in materia di sicurezza, riservatezza e accesso ai dati.

L'AgID ha già diffuso le linee guida per il FSE con le specifiche tecniche per l'interoperabilità dell'infrastruttura ed ha istituito un portale dedicato, con valutazione e approvazione dei piani di tutti i piani di progetto regionali.

Nel settembre 2015 è stato firmato il regolamento del FSE, nel Decreto del Consiglio dei Ministri, da parte del Ministro della Salute; attualmente le Regioni italiane presentano un'elevata variabilità in termini di infrastrutture, architetture applicative, di modalità di informatizzazione del FSE, nonché dei documenti resi disponibili al cittadino. Le uniche due Regioni ad aver istituito ed avviato il FSE sono ad oggi l'Emilia-Romagna con il progetto "SOLE" e il Trentino con "TreC"; in stato

avanzato la Toscana e la Lombardia, con alcune criticità da affrontare, mentre Piemonte, Veneto, Friuli-Venezia Giulia, Puglia e Sardegna hanno reso operative alcune componenti del FSE Fascicolo. Le restanti Regioni non dispongono ancora del FSE. Entro la scadenza del 31 dicembre 2015 avrebbe dovuto essere attivato un sistema unificato a livello nazionale per l'interoperabilità tra infrastrutture regionali⁸⁰.

In generale la digitalizzazione dei processi sanitari dovrebbe indurre il miglioramento dei flussi operativi e del processo diagnostico assistenziale, per riduzione dei tempi e dei costi necessari al completamento di un processo diagnostico. La diffusione e l'implementazione del FSE potrebbero favorire questo miglioramento e rendere possibile anche la costruzione di sistemi monitoraggio a supporto di programmazione, gestione, controllo e valutazione dell'assistenza sanitaria, accessibili in modo strutturato anche per finalità di governo, nonché supportare studi e ricerche scientifiche in campo medico, biomedico ed epidemiologico. Per la realizzazione concreta di tale progetto si rende tuttavia necessario un intervento normativo completo a livello nazionale, che definisca i presupposti per il trattamento dei dati per queste finalità nei contesti regionali e nell'ambito del Nuovo Sistema Informativo Sanitario al livello centralizzato.

Anche nel contesto internazionale è documentata la letteratura l'utilità del FSE, cosiddetto "Electronic Health Record" (EHR) o "Personal Health Record" (PHR) a supporto del processo diagnostico-assistenziale e terapeutico in ambito sanitario, per la misurazione ed il monitoraggio della qualità delle cure in un determinato paese e per la ricerca scientifica traslazionale⁸¹.

Diversi modelli di EHR sono stati sviluppati in alcuni paesi europei⁸². In Inghilterra sono stati messi in atto numerosi tentativi per la creazione di un'infrastruttura di informazione sanitaria a livello del sistema sanitario pubblico, con notevole dispendio economico per modernizzare i sistemi di raccolta, archiviazione e utilizzo dei dati sanitari. La progettazione e la realizzazione ancora di EHR sono in fase di sperimentazione, come parte di un programma nazionale per la tecnologia nell'informazione sanitaria.

In Germania è in corso di programmazione un modello di EHR sotto forma di tessera sanitaria smart-card elettronica che rende possibile l'elaborazione di applicazioni di assicurazione sanitaria e trascrizioni elettroniche di dati sanitari.

La Francia, in ambito sanitario a livello nazionale, ha avviato fin dal 1998 l'utilizzo sicuro di sistemi elettronici, che includevano una client smart card per dati sanitari e socio-sanitari da rendere disponibili in rete agli operatori in presenza dei pazienti. Nel 2004, il Parlamento francese ha approvato una legge per l'attuazione PHRs elettronici, cui ha fatto seguito la creazione nel 2007 di un dossier sanitario elettronico personale per ogni cittadino. Fatta eccezione per Scozia, Svezia e Danimarca, dove sono attivi alcuni modelli di EHR per limitati bacini di popolazione o per l'utilizzo d'informazioni digitalizzate protette da parte dei medici di medicina generale, in letteratura non è documentata l'esistenza di modelli di EHR in altri paesi europei.

In ambito extra-europeo la letteratura individua l'Australia e la Nuova Zelanda come paesi all'avanguardia per l'attuazione di piani e la creazione d'infrastrutture per le tecnologie di informazione sanitaria, anche se in nuova Zelanda la realizzazione di EHR personali non rientra nelle priorità delle strategie di governo. In Australia è attualmente attivo un sistema informatizzato in rete, che permette agli utenti iscritti di creare e gestire il dossier sintetico individuale dei dati sanitari.

In America, gli Stati Uniti e il Canada sono gli unici paesi che hanno avviato strategie operative efficaci per la creazione e l'implementazione dell'EHR e che dispongono d'infrastrutture definite e di adeguati sistemi di supporto per l'attuazione di tali strategie.

In Africa non vi sono stati in cui sia documentato un progetto di creazione di EHR.

A Taiwan, il Dipartimento di Salute sta sviluppando fin dal 2001 il National Electronic Personal Health Record al fine di promuovere la qualità dei servizi sanitari, avviando numerosi progetti tra cui l'utilizzo di smart card per gestione delle pratiche di assicurazione sanitaria e l'adozione di norme internazionali per la gestione dell'informazione sanitaria. Dopo anni di lavoro, questo progetto ha tuttavia ancora potenzialità ed applicazioni limitate.

In generale lo scenario internazionale rivela che in ambito di FSE individuale si è ancora in una fase embrionale e sono numerose le sfide ancora da affrontare, soprattutto per i paesi in via di sviluppo, per la programmazione e l'attuazione di questo sistema tecnologico di gestione dell'informazione sanitaria, che sembra subire l'influenza, in diversi paesi, di differenti fattori ambientali, tra cui le caratteristiche operative, organizzative e culturali e la prospettiva di sviluppo sociale e tecnologico in ciascun paese. Nella maggior parte dei paesi vi è una sostanziale carenza di norme specifiche che regolamentino l'accesso e la libera circolazione di dati sanitari sul web, con conseguenti grosse questioni di affidabilità dei sistemi predisposti, di gestione della privacy e di riservatezza dei dati personali⁸³, mentre altre problematiche sembrano maggiormente connesse ai bassi livelli di alfabetizzazione informatica e d'informazione in ambito di salute pubblica presenti in molti paesi in via di sviluppo.

1.5 I MALATI RARI SENZA DIAGNOSI

1.5.1 Scenario di riferimento

Le malattie rare costituiscono un grande problema di salute pubblica, perché sono spesso ad andamento cronico, sono progressive, degenerative, associate a grave disabilità fisica, cognitiva, neurologica, cronicamente debilitanti ed invalidanti e potenzialmente fatali. Per l'elevato grado di complessità genotipica e fenotipica, come visto, sono frequentemente difficili da inquadrare e trattare: i malati rari incontrano spesso difficoltà peculiari in termini di accessibilità a cure di elevata qualità, in primis a causa della generale carenza di conoscenze medico-scientifiche e di conseguenza a causa di una serie di complesse problematiche connesse al processo d'inquadramento diagnostico⁸⁴. La difficoltà nel giungere ad una diagnosi corretta, che in molti casi è un processo che richiede anni, talora decenni, configurano un drammatico ostacolo per i malati rari e possono comportare conseguenze drammatiche in termini di qualità di vita. La possibilità di ottenere un corretto inquadramento diagnostico implica un notevole dispendio economico e di energie per i pazienti stessi e per le loro famiglie, che devono sottoporsi a numerose visite specialistiche e complessi esami diagnostici, spesso in mancanza di un reale supporto da parte delle strutture del sistema sanitario ed in assenza di centri di riferimento multidisciplinari d'eccellenza, di servizi ed operatori sanitari qualificati, di trattamenti adeguati e specifici sotto il profilo eziologico, ma con necessità di rivolgersi a molteplici differenti strutture, frequentemente lontane e disperse

sul territorio. In questo contesto la comunicazione diagnostica è spesso inadeguata e la diagnosi stessa risulta non corretta, con conseguenze individuali anche molto gravi per il malato raro, che vanno dall'aumentato rischio di complicanze e sequele tardive e dal decadimento delle condizioni psico-fisiche e cliniche, al profondo disagio psicologico in alcuni casi fin anche alla morte.

I ritardi nel processo d'inquadramento diagnostico possono variare molto a seconda del tipo di malattia rara in questione, con intervalli temporali che possono andare da un mese e mezzo fino a trent'anni dalla comparsa dei primi sintomi al momento della corretta diagnosi; vi è comunque grande differenza anche all'interno dello stesso gruppo di malattie, con ritardi diagnostici che possono variare, per una stessa malattia, dai quindici ai trent'anni dalla comparsa dei primi sintomi.

Gli errori diagnostici sono molto frequenti, con circa il 40% di pazienti che riceve almeno una diagnosi incorretta prima di ottenere il giusto inquadramento diagnostico; tale aspetto implica inevitabilmente la messa in atto di trattamenti non adeguati, che includono interventi chirurgici e altri trattamenti invasivi, o, spesso, inadeguati percorsi di consultazione psicologica o di ospedalizzazione e medicalizzazione psichiatrica. L'errore diagnostico causa inoltre l'interruzione del processo d'inquadramento diagnostico, con conseguente ulteriore ritardo nel raggiungimento della correttezza diagnostica.

In molti casi sono le risorse personali ed economiche dei pazienti stessi e delle loro famiglie a consentire la prosecuzione del percorso d'inquadramento diagnostico ed il raggiungimento della diagnosi, con la personale identificazione del sospetto diagnostico ed l'individuazione dei laboratori diagnostici o dei centri altamente specializzati in esami di conferma diagnostica. Per quasi tutti questi casi di autogestione del processo diagnostico i ritardi risultano notevolmente maggiori, rispetto a quei casi invece in cui tale processo è sostenuto dall'intervento di risorse professionistiche in ambito sanitario.

Le conseguenze dei ritardi e degli errori diagnostici riguardano soprattutto la difficoltà per i pazienti nell'ottenere adeguati trattamenti medici, con aggravamento delle condizioni cliniche e peggioramento dello stato di salute in generale; altre gravi conseguenze, come già accennato, riguardano il decadimento psicologico, cognitivo ed

emotivo, talora la morte dei pazienti, oppure, in ambito psico-sociale la nascita di altri individui affetti nelle famiglie di malati rari, il disadattamento delle famiglie stesse e la perdita generale di fiducia nelle infrastrutture medico sanitarie.

In generale quindi i malati rari senza diagnosi, ovvero pazienti che presentano segni e sintomi verosimilmente attribuibili ad una malattia rara, ma la cui diagnosi non è stata ancora formulata è un gruppo eterogeneo, che comprende pazienti che hanno subito ritardi diagnostici, errori diagnostici (ovvero affetti da malattie note non riconosciute) ed infine i pazienti con nuove entità nosologie sconosciute; questi pazienti rappresentano una sfida su molti fronti per i sistemi sanitari: da un lato l'utilizzo eccessivo di percorsi diagnostico-terapeutici inadeguati, dall'altro la persistenza di bisogni assistenziali e terapeutici non soddisfatti. I ritardi diagnostici possono compromettere il percorso terapeutico dei pazienti e familiari, con il risultato di elevati costi umani e sociali e gli errori diagnostici possono interrompere e fuorviare un corretto percorso diagnostico e portare a prescrizioni e trattamenti inappropriati, dispendiosi, talora addirittura nocivi.

Ad oggi il problema dei pazienti senza diagnosi è stato universalmente riconosciuto a livello mondiale come centrale nella programmazione sanitaria della rete di assistenza per malati rari, ma appare ancora irrisolto: nonostante la centralità rilevante della questione a tutt'oggi non sono ancora state identificate le politiche sanitarie migliori e le conseguenti strategie programmatiche.

1.5.2 Undiagnosed Diseases Program in USA

Fonti dell'Office of Rare Diseases Research (ORDR) dell'NIH, indicano che il 33% dei malati rari ottiene un corretto inquadramento diagnostico solo dopo un intervallo di tempo lungo, variabile tra uno e cinque anni della comparsa di primi sintomi di malattia, mentre il 15% deve attendere ben oltre i cinque anni⁸⁵.

In USA sono stati attuati tentativi rilevanti in tema di malati rari senza diagnosi: in particolare, dal maggio 2008, è attivo Undiagnosed Diseases Program (UDP), in collaborazione con National Institute of Health (NIH) Clinical Center, che prevede l'integrazione dell'informazione genetica ottenuta da nuove tecniche d'indagine, con dati clinici noti e con il profilo fenotipico di pazienti senza diagnosi. L'UDP è stato istituito con l'obiettivo principale di far fronte alle esigenze di pazienti e famiglie con

malattie rare multisistemiche, rimaste non diagnosticate anche dopo un lungo ed esaustivo percorso di valutazione diagnostica⁸⁶. Il programma garantisce ai malati rari senza diagnosi la possibilità di sottoporsi ad innovative risorse diagnostiche, quali le nuove tecniche di sequenziamento genomico, in grado, potenzialmente, di identificare nuove entità nosologie e di comprendere al meglio le basi biologiche delle malattie e la fisiologia umana⁸⁷.

La popolarità di UDP è cresciuta rapidamente ed il suo sviluppo è il risultato di uno sforzo di sensibilizzazione clinica, che ha coinciso con la rapida espansione delle tecniche “Whole exome sequencing” (WES) e “Whole genome sequencig” (WGS).

Il programma è iniziato con la registrazione, previo consenso informato, dei pazienti afferenti all’NIH Clinical Center (NIH-CC), nell’ambito del protocollo di ricerca “Diagnosis and Treatment of Patients with Inborn Errors of Metabolism and Other Genetic Disorders”. La struttura sanitaria ed il medico di riferimento abituale, che hanno continuato ad occuparsi della continuità assistenziale-terapeutica del paziente, ne hanno fornito un scheda riassuntiva e la cartella clinica, completa di referti d’imaging e campioni biologici (vetrini istologici, materiale bioetico). Il paziente è stato quindi registrato all’interno di un database clinico e le informazioni della cartella clinica sottoposte a revisione, allo scopo di stabilire l’idoneità all’arruolamento nello studio, selezionando in particolare i casi affetti da patologie a probabile eziologia genetica, e di delineare il profilo fenotipico del paziente; molti specialisti clinici, appartenenti a molteplici discipline (dermatologia, patologia generale, oculistica, radiologia, virologia, ematologia, cardiologia, pneumologia, reumatologia, medicina di laboratorio, immunologia, malattie infettive, endocrinologia, epatologia, nefrologia, odontoiatria, farmacologia, psichiatria, neurologia, terapia del dolore e cure palliative, medicina interna, pediatria, genetica, fisioterapia, disturbi del metabolismo osseo, e otorinolaringoiatria) si sono contemporaneamente impegnati nella revisione delle cartelle cliniche. I pazienti arruolati nel programma sono stati tratti in degenza una settimana presso l’NIH-CC per completamento diagnostico ed un team di circa 60 operatori sanitari ha presieduto, nell’ambito di meeting periodici, all’aggiornamento dati ed alla discussione dei casi clinici.

Il DNA per analisi genomica, estratto da sangue periferico, è stato inviato presso il National Human Genome Research Institute (NHGRI) per l'analisi SNP, ovvero l'identificazione di polimorfismi specifici a singolo nucleotide in ciascun campione di DNA, e presso l'Intramural Sequencing Center dell'HIC per l'applicazione di tecniche WES e WGS sul DNA del probando e dei familiari con caratteristiche geneticamente rilevanti.

Le informazioni ottenute sono state sottoposte a successiva elaborazione, con identificazione ove possibile delle varianti a significato patogeno, e di ciascuna variante genomica comune attraverso il confronto con i dati contenuti in database SNP di riferimento internazionale.

Nei primi tre anni circa di attività UDP ha ricevuto più di 4700 candidati, ha selezionato e sottoposto a revisione 1191 cartelle cliniche, da cui sono stati selezionati 272 casi, di questi il 20-25% ha ricevuto una diagnosi definitiva, nella maggior parte dei casi di malattie note, tra cui 25 malattie rare o ultra rare. Sono state inoltre identificate 2 nuove entità nosologiche, mai descritte in precedenza, ed ampliata la conoscenza sul profilo fenotipico di alcune malattie note.

I risultati ottenuti complessivamente in 6 anni di lavoro sono apparsi controversi; da un lato l'UDP ha dimostrato un ampio valore nell'implementazione di una metodologia per il corretto inquadramento di pazienti affetti da malattia rara senza diagnosi e ha sottolineato il ruolo fondamentale delle tecniche di sequenziamento genomico in questo processo, soprattutto in relazione a malattie molto complesse sotto il profilo fenotipo-genotipico. D'altro canto le risorse impiegate appaiono mastodontiche, a partire dalla complessa e indaginosa fase di arruolamento casi, fino al dispendioso processo di revisione delle cartelle e dei casi clinici, spesso con coinvolgimento di numerosi specialisti e diversi centri di riferimento, con notevole dispendio in termini di tempo e costi e quindi con notevoli problematiche di sostenibilità operativa nella pratica clinica. Attualmente infatti il programma è fermo e ne è messa in dubbio la futura sostenibilità.

L'UDP ha affrontato per la prima volta una questione irrisolta, ovvero l'inquadramento diagnostico di pazienti affetti da complesse patologie multisistemiche e può servire come modello per l'applicazione clinica delle nuove tecnologie di

sequenziamento genomico, ma sembra emergere la necessità di sviluppare sistemi più efficienti di utilizzo della nuova informazione genotipica ad integrazione del profilo fenotipico ed un modello sanitario che garantisca un arruolamento dei casi efficace e realizzabile nella pratica clinica. Diversi sistemi automatizzati quali l'analisi di immagini digitali delle caratteristiche facciali e gli Adaptive Inference Systems Fuzzy hanno dato risultati promettenti, ma solo su limitati campioni di popolazione⁸⁶.

1.5.3 Gli scenari di ricerca

A livello nazionale è crescente l'interesse della comunità medico-scientifica per la creazione di modelli e sistemi di raccolta dell'informazione clinica, sostenuta dall'utilizzo di moderne tecniche di sequenziamento genomico, di pari passo con lo sviluppo di registri e reti assistenziali regionali e nazionali specifiche per i malati rari; a tal proposito Centri d'Eccellenza e riferimento per diagnosi e cura di specifiche malattie rare o gruppi di malattie rare correlate in alcune regioni italiane, hanno istituito progetti mirati all'implementazione delle conoscenze in ambito di malattie rare con caratterizzazione della corrispondenza fenotipo-genotipica, ed al sostegno del processo diagnostico-assistenziale dei malati rari, con particolare riguardo ai pazienti senza diagnosi o con un generico sospetto diagnostico.

In questa sezione si fa riferimento al Progetto di ricerca Regione-Università "Next Generation Sequencing and Gene Therapy to Diagnose and Cure Rare Diseases in Regione Emilia Romagna (RARER)" ed al Progetto Ministeriale "A multicenter collaborative research network for the identification and study of rare undiagnosed patients: the impact on the rare disease National Health Service network (UnRareNet)".

Next Generation Sequencing and Gene Therapy to Diagnose and Cure Rare Diseases in Regione Emilia Romagna (RARER)

Il programma di ricerca Regione-Università RARER, approvato e finanziato dalla Regione Emilia Romagna e presentato dall'Azienda Ospedaliera Universitaria di Modena, si è posto come obiettivo generale il miglioramento nella gestione assistenziale dei pazienti con malattia rara, attraverso la creazione di registri, definiti nei contenuti ed orientati a specifiche malattie rare o gruppi di malattie rare correlate, idonei a raccogliere in modo omogeneo ed esaustivo dati riguardanti le caratteristiche fenotipiche, genotipiche e dei percorsi assistenziali di pazienti afferenti ai Centri Clinici

di riferimento per patologia in Emilia Romagna. Tali registri sono stati progettati per rendere utilizzabile l'informazione clinica per identificare i determinanti genetici e ambientali dell'espressività clinica, la storia naturale della patologia e la risposta al trattamento.

La Regione Emilia Romagna dispone da tempo di un sistema informativo specifico per le malattie rare, che consente il censimento dei malati rari presenti e seguiti in Regione e la raccolta dell'informazione clinica-assistenziale a sostegno della pratica clinica; gli aspetti strategici del progetto RARER sono stati:

- l'implementazione di registri dettagliati come parte estensiva degli applicativi già a regime, per rendere omogeneo e fruibile da tutti gli attori, centri di ricerca e governo regionale, il patrimonio informativo raccolto, con possibilità di estendere eventualmente il modello ad altre tipologie di malattie rare attualmente non comprese nello studio e di facilitare il lavoro di clinici e ricercatori attraverso l'utilizzo di un'unica piattaforma informativa;
- lo sviluppo strutturale dei registri sulla base di una logica clinica gerarchica suddivisa per organi ed apparati, che sotto-definisca i contenuti clinici a differenti e sempre più dettagliati livelli, per descrizione delle singole entità cliniche attraverso sistemi di codifica specifici ed inter-operabili, che consentano la classificazione clinica corretta delle malattie note, ma siano potenzialmente utili anche per entità cliniche sconosciute similari o fenotipicamente prossime;
- l'utilizzo di nuove tecniche di sequenziamento genomico NGS, per fornire ai pazienti una diagnosi genotipica accurata, completa e a basso costo, attraverso l'identificazione personalizzata della corrispondenza fenotipo-genotipica, per identificare eventualmente le basi genetiche di alcune malattie ad ereditarietà mendeliana e sviluppare protocolli di terapia genica.

Il sistema di compilazione dei registri di malattia è stato supportato da un modello unico e comune di database relazionale, strutturato in modo da contenere l'informazione clinica organizzata secondo moduli differenti (modulo anagrafica,

modulo diagnosi clinica, modulo prescrizioni, modulo diagnosi genomica etc.) e descritta all'interno di categorie predefinite (segni e sintomi, comorbidità, disabilità, danni funzionali etc); il sistema prevede l'archiviazione dell'informazione clinica del paziente al primo e ad ogni successivo accesso al centro, in modo tale da consentire la gestione degli episodi clinici ed il follow up per l'intero ciclo di vita del paziente. La classificazione delle entità fenotipiche è basata su un approccio standardizzato secondo le linee guida nazionali ed europee, che integra la struttura monodimensionale delle classificazioni ICD9-CM e ICD10, ai codici MIM ed alla struttura multigerarchica ad albero di Orphanet. Anche l'informazione genomica è raccolta secondo nomenclature standardizzate e validate a livello internazionale.

La compilazione degli 8 registri di malattia è stata gestita a livello locale, a cura di centri accreditati, di riferimento per 8 malattie rare o gruppi di malattie rare correlate con sede in Regione Emilia Romagna, nelle città di Modena, Reggio Emilia, Parma e Bologna: Emocromatosi Ereditaria, Degenerazione Retinica Ereditaria; Osteocondromatosi Multipla; Distrofia Muscolare di Duchenne; Distrofia Muscolare Facio-scapolo-omerale; Ipoacusie Ereditarie; Neurofibromatosi; Ipercolesterolemie Familiari.

I registri, attualmente a regime, sono basati su un'interfaccia web compatibile al modello dei registri regionali e nazionali; la gestione in rete è presieduta da un'infrastruttura centralizzata, che prevede anche l'integrazione dei registri nel sistema informativo regionale specifico per i malati rari in Emilia Romagna e nel Registro Regionale per le malattie rare del Veneto. La Regione Emilia Romagna mantiene a proprio carico tutte le questioni connesse alla sostenibilità a lungo termine del sistema.

A multicenter collaborative research network for the identification and study of rare undiagnosed patients: the impact on the rare disease National Health Service network (UnRareNet)

Nel Sistema Sanitario Nazionale l'accesso alle prestazioni, esclusa la fase pre-diagnostica, è per un malato raro, garantita dalla formulazione di una diagnosi definitiva; un esempio applicativo è il sistema informativo V-ER-L RDIS, attivo in Veneto, Emilia Romagna e Liguria per la gestione delle cartelle cliniche ed il monitoraggio epidemiologico di pazienti con diagnosi malattia rara, che possono

beneficiare di trattamenti necessari in regime di gratuità, solo se prescritti attraverso il sistema informativo stesso. Per i pazienti senza diagnosi le difficoltà in termini di accessibilità alle cure sono in tal senso enormi, come pure di opportunità ad essere reclutati in studi clinici, programmi di ricerca o essere iscritti a registri di patologia. Un programma specificamente rivolto a queste questioni ha un importante valore aggiuntivo se si considerano i principi di non discriminazione, anche in ambito di malattie rare.

Il problema dei malati rari senza diagnosi è inoltre molto rilevante in alcune aree del Paese, data l'elevata variabilità di organizzazione e qualità dei servizi sanitari offerti a livello regionale, che determinano continue migrazioni dei pazienti per ottenere la diagnosi e gli appropriati precorsi di cura; i più recenti Piani Sanitari Nazionali hanno riconosciuto l'esistenza di differenze interregionali nella disponibilità e nella qualità dei servizi forniti, quali ad esempio procedure diagnostiche moderne ed innovative, che non rientrano nei cosiddetti livelli essenziali di assistenza, come aspetto molto critico ed in ambito di malattie rare per i pazienti non diagnosticati, la mobilità inter-regionale ed anche estera è particolarmente alta. Ciò implica oneri aggiuntivi per pazienti e famiglie, in termini di costi umani e sociali e costi elevati per il sistema sanitario, in termini di oneri finanziari sulle regioni con bassa organizzazione di servizi e conseguente alto tasso di mobilità. La creazione di una rete di collaborazione e di una piattaforma ad uso comune tra centri competenti e professionisti rivolti a diagnosi e cura di malattie rare, che possano ridurre l'utilizzo inopportuno ed eccessivo dei servizi di assistenza sanitaria con work-up diagnostici inutili e trattamenti non adeguati, è di grande rilevanza ed interesse pubblico e particolarmente utile all'implementazione dei percorsi diagnostici complessi dei pazienti affetti da malattia rara. La creazione di una banca centralizzata di informazioni cliniche e di campioni biologici di malati rari senza diagnosi potrebbe inoltre costituire un'enorme risorsa in ambito di ricerca clinica.

In questo contesto generale rivolto alla progettazione di una rete assistenziale italiana per i malati rari, si inserisce il progetto UnRareNet, che coinvolge i Centri Accreditati per malattia rara compresi nel sistema V-ER-L RDIS, ospedali e strutture sanitarie locali, ed altri Centri, in diverse Regioni italiane, riconosciuti per grande competenza in diagnosi e cura di specifiche malattie rare o gruppi di malattie rare, situati in aree geografiche d'interesse per la presenza di pazienti non diagnosticati.

Il progetto prende vita dall'esperienza decennale delle reti regionali per i malati rari e si costituisce nell'ambito della Ricerca Finalizzata finanziata dal Ministero della Salute e dal Ministero dell'Università e della Ricerca, in collaborazione con l'Ospedale Pediatrico Bambino Gesù, quale centro di sperimentazione principale e con i seguenti centri collaboratori: Dipartimento di Pediatria Azienda Ospedaliera Federico II Napoli, UOC Neurogenetica Molecolare Istituto Neurologico Carlo Besta Milano, Dipartimento ligure Genetica EO Galliera Genova, Istituto Gaslini Genova, Istituto Dermopatico dell'Immacolata-IDI-IRCCS Roma, Rete Bio-banche Telethon (TGBN).

Il Centro Coordinamento Malattie Rare della Regione Veneto dispone già di un sistema informativo contenente una raccolta sistematica di segni e sintomi correlati a malattie rare e si occupa della gestione dei percorsi assistenziali dei malati rari con diagnosi sospetta o certa.

Nell'ambito del progetto UnRareNet si prevede una definizione più dettagliata di questi contenuti, integrandoli, se necessario, con i dati di Orphanet, Oxford Dismorphology Database, OMIM ed altre fonti internazionali, in relazione a diverse, eterogenee e non comuni caratteristiche che possono essere tipiche delle condizioni patologiche non diagnosticate; in accordo alle linee guida europee su codificazione e classificazione delle malati rare è stata inoltre proposta una classificazione più sistematica delle entità rare, con attribuzione di codici internazionali, relativi a diversi sistemi di classificazione ed approcci logici a più di 6.000 entità nosologiche. Particolare attenzione è stata posta alle malattie rare multi-sistemiche, per la maggior parte difficili da diagnosticare e classificare a livello nosologico. Pre-requisiti di sviluppo del progetto sono la disponibilità di competenze e strumenti che consentano l'utilizzo appropriato di moderne tecniche diagnostiche, come l'array CGH (Comparative Genomic Hybridization) e la Next Generation Sequencing (NGS), per identificazione di nuove varianti genomiche, ed il coinvolgimento ufficiale di grandi associazioni, come la Federazione Italiana delle Associazioni di pazienti affetti da malattia rara "Uniamo" che può contribuire alla diffusione delle informazioni, non esistendo al momento nessuna specifica associazione che rappresenti i malati rari senza diagnosi.

Gli obiettivi del progetto sono:

- 1) la costruzione di una piattaforma web nazionale, che colleghi i Centri di competenza ed i servizi sanitari attraverso una rete privata e dedicata, per la gestione di una banca dati clinica connessa ad una repository di campioni biologici di pazienti senza diagnosi probabilmente affetti da malattia rara; per la realizzazione di questo obiettivo devono essere definiti i contenuti clinici, le modalità di registrazione e di elaborazione dei quadri clinici, degli esiti di esami di laboratorio e di altri esami, la metodologia standard per la raccolta di imaging e di materiale fotografico, per l'aggregazione e per la descrizione dei dati clinici secondo sistemi di classificazione specifici, un protocollo di regole per l'accesso e l'utilizzo dei dati e dei campioni contenuti nella biobanca, stabilito in collaborazione con UNIAMO e validato dal un Comitato Etico ed infine una metodologia statistica per l'elaborazione dati;
- 2) l'identificazione, tra tutti i casi registrati, di quelli ad elevato sospetto diagnostico, attraverso un sistema di screening che assegni in modo automatico una diagnosi probabilistica, attraverso l'analisi automatizzata delle caratteristiche fenotipiche e dimorfologiche di ciascun paziente ed il confronto con profili clinici noti, selezionati per diverse entità nosologiche o con dimorfismi sindromici noti, contenuti in database dimorfologici già esistenti. Per i casi così selezionati è previsto il re-invio presso il Centro Accreditato per competenza, se possibile per la conferma diagnostica.
- 3) l'individuazione, tra i casi rimasti ancora senza diagnosi dopo questo processo di attribuzione diagnostica, di sottogruppi omogenei di pazienti attraverso metodi di clusterizzazione basati sull'omogeneità fenotipica, per la successiva esecuzione di test genetici mirati a identificazione di varianti di DNA comuni con probabile relazione di causalità o di nuove varianti genomiche responsabili di entità nosologiche non conosciute.

Il progetto proposto presenta molti aspetti rilevanti per il SSN, con potenziale impatto positivo sulla qualità di vita dei malati rari ed è appropriato rispetto ad obiettivi importanti indicati nei Piani Sanitari Nazionali in termini di miglioramento dell'accessibilità alle risorse diagnostiche e della qualità delle cure e di riduzione della disabilità a lungo termine; questi aspetti sono cruciali nel caso dei pazienti senza

diagnosi, che come detto si trovano ad affrontare lunghi ed infruttuosi percorsi assistenziali con conseguenti perdita di fiducia nel sistema sanitario, senso di isolamento, frustrazione, gravi complicanze e sequele tardive.

1.5.4 I problemi comuni

Modelli e sistemi di raccolta dell'informazione clinica in ambito di malattie rare, oggi potenzialmente disponibile nella sua complessa corrispondenza fenotipo-genotipica grazie a moderne tecniche d'indagine genomica, così come la realizzazione di sistemi di codifica e classificazione delle malattie rare, validati e condivisi a livello internazionale, costituiscono una riconosciuta risorsa per giungere a nuova e più approfondita conoscenza sull'eziologia, la storia naturale queste malattie e la risposta ai trattamenti e per disporre di misure epidemiologiche e d'impatto sulla popolazione in termini di salute e qualità di vita, soprattutto per quei pazienti, ancora senza diagnosi, per i quali l'accessibilità ad adeguati percorsi diagnostico-terapeutici appare, come visto, gravemente compromessa. In tal senso negli ultimi tempi si è assistito alla diffusione massiccia di numerosi ed eterogenei database e registri di malattia e di programmi e modelli di raccolta dell'informazione clinica la cui validità, pur trattandosi di strumenti molto promettenti, rappresenta una sfida importante ed attuale per la comunità scientifica date alcune problematiche comuni ancora irrisolte.

La prima questione riguarda la scarsissima flessibilità metodologica e la grande variabilità ed eterogeneità di questi strumenti in termini di struttura, contenuti, terminologia di riferimento, obiettivi e logica classificatoria, caratteristiche che al momento ne compromettono l'utilizzo trasversale, al di fuori dello scopo per cui sono stati direttamente pensati, e ne mettono in discussione la validità come modelli interoperabili ed universali per la raccolta dell'informazione clinica in ambito di malattie rare.

Altra questione riguarda la grande mole e complessità numerica e strutturale delle nuove informazioni genomiche, acquisite grazie alle moderne tecniche di sequenziamento del DNA, che pongono di fronte a notevoli problematiche metodologiche in termini di codifica, raccolta, interpretazione e classificazione; come visto non esistono ad oggi modelli o sistemi di riferimento la cui validità sia universalmente riconosciuta e che garantiscano operativamente il conseguimento di

conoscenze necessarie a comprendere il reale impatto della variabilità genetica umana in termini di salute. La correlazione significativa tra varianti genomiche e profilo fenotipico delle malattie è in moltissimi casi ancora oscura e di difficile interpretazione, data l'enorme complessità clinica ed eziopatogenetica che caratterizza la stragrande maggioranza delle malattie rare multisistemiche ad ereditarietà mendeliana.

I sistemi descritti di raccolta dell'informazione clinica, completa nella sua corrispondenza fenotipo-genotipica, come detto, hanno dato risultati promettenti perché si sono dimostrati potenzialmente in grado di fornire misure reali, basate su dati di popolazione, dell'impatto di queste malattie sulle salute umana, ma sono apparsi ancora scarsamente sostenibili nella pratica clinica, perché frequentemente frutto d'iniziative di singole istituzioni o protocolli di ricerca, con limitata copertura geografica, altamente dispendiosi in termini di risorse economiche, di tempo e mano d'opera e di ancora limitato valore epidemiologico, proprio perché gravati da caratteristiche di incompatibilità e scarsa qualità metodologica. Tali caratteristiche ne mettono in discussione il loro valore anche in ambito di ricerca medica, per cui è necessario disporre di casistiche sufficientemente ampie ed adatte a studi clinici sui meccanismi eziopatogenetici del danno ed a studi di sperimentazione clinica.

La questione resta aperta e di cruciale interesse per la comunità scientifica, che si è posta come obiettivo la ricerca di soluzioni metodologiche che consentano di raccogliere l'informazione necessaria attraverso strumenti generali, comuni ed interoperabili, che garantiscano il raggiungimento di una conoscenza globale ed olistica dell'impatto delle malattie rare sulle salute umana e che possano quindi orientare le decisioni politiche e le conseguenti strategie di programmazione sanitaria.

2 OBIETTIVI

2.1 Creazione di strumenti di descrizione analitica e precodificata dell'insieme di diversi fenotipi e genotipi

In regione Veneto è in uso dal 2002 un sistema informativo specifico, costituito da un registro di pazienti con diagnosi di Malattia Rara e da un sistema di monitoraggio assistenziale-terapeutico, che gestisce i percorsi di cura dei singoli pazienti. Il sistema è gestito dal Centro Coordinamento Regionale per le Malattie Rare del Veneto, all'interno di una rete assistenziale intranet, che provvede al coordinamento dei presidi accreditati in rete (Area Vasta: Veneto, Friuli Venezia Giulia, provincie autonome di Trento e Bolzano), allo scambio informativo con altri centri regionali italiani, alla consulenza e al supporto ai Medici del Servizio Sanitario Nazionale, alla collaborazione ad attività formative per operatori sociali ed associazioni di famiglie e volontari, alla diffusione d'informazioni al cittadino. Il sistema contiene una banca dati clinica con sezione dedicata a segni e sintomi principali presenti in ciascuna patologia.

Date le premesse culturali introduttive si è ritenuto interessante sviluppare un progetto di ricerca con l'obiettivo primario di realizzare uno strumento unico di raccolta di dati analitici descrittivi il fenotipo e il genotipo di pazienti con diagnosi nota o sospetta, all'interno di gruppi di malattie rare selezionate, ovvero di ridefinizione genomica e riclassificazione delle entità nosologiche, con codifica delle varianti cliniche e dei determinanti genetici specifici del fenotipo clinico.

Lo sviluppo dello strumento, basato su una piattaforma informatizzata, si realizza a partire dal modello del Registro Regionale per le Malattie Rare già in uso in Regione Veneto e prevede l'implementazione della banca dati clinica, con organizzazione delle sezioni secondo struttura gerarchica suddivisa per organi ed apparati, e la realizzazione di una banca dati di varianti genomiche; l'integrazione delle banche dati configura uno strumento unico di descrizione dettagliata del fenotipo e del genotipo. Tale strumento deve svilupparsi in accordo ad una logica basata sull'approccio clinico, ma adattarsi anche ad un contesto di ricerca, deve poter descrivere l'informazione clinica al maggior grado di completezza teoricamente possibile e avere caratteristiche di flessibilità ed adattabilità a situazioni molto diverse, oltre a capacità analitiche non predefinite, cioè non ancorate a conoscenze già acquisite

su specifiche entità nosologiche, ma aperto all'acquisizione di nuove informazioni non note e quindi completamente elaborabile, differenziandosi sostanzialmente dai vari registri specifici di patologia esistenti a livello nazionale e internazionale, per l'assoluta generalizzazione e possibile utilizzo per una qualsiasi delle malattie note di cui i soggetti arruolati potrebbero essere considerati affetti.

L'obiettivo è quindi la creazione di un database relazionale in grado di gestire la raccolta aperta d'informazioni cliniche precodificate di pazienti con diagnosi nota o sospetta di malattia rara, al maggior grado di completezza e dettaglio possibile, subito utilizzabili per procedure elaborative complesse; l'infrastruttura tecnologica di base deve garantire accessibilità ventiquattrore su ventiquattro, per trecentosessantacinque giorni all'anno, essere in grado di gestire l'utenza ed i dati in modo tracciabile, in sicurezza e privacy e rispondere all'ulteriore obiettivo di integrare l'informazione clinica e genetica (anche derivante dall'applicazione di tecniche di Next Generation Sequencing).

2.2 Trasferimento degli strumenti descrittivi in prodotti utili per la gestione clinica

Considerate le premesse introduttive sui pazienti affetti da malattie rare senza diagnosi definita ed i tentativi fatti in passato in particolare con l'UDP, è evidente la difficile sostenibilità di un sistema che coinvolga una rete di centri qualificati dove inviare prioritariamente i malati con diagnosi non definita. È emersa quindi la necessità di sviluppare un modello sanitario che garantisca un arruolamento dei casi efficace e realizzabile nella pratica clinica.

Il progetto di dottorato si pone come obiettivo ulteriore la creazione di un prodotto specifico per alcune necessità attualmente emergenti in ambito di malattie rare, soprattutto rispetto ad efficienza, rapidità e correttezza d'inquadramento diagnostico, attraverso il trasferimento dello strumento descrittivo, realizzato per raccolta dell'informazione clinica fenotipica e genotipica di pazienti con malattie o gruppi di malattie rare selezionate, in prodotto utile alla gestione clinica, ovvero nel modulo generalizzato di Cartella Clinica Informatizzata (CCI). La CCI rappresenta il dossier contenente l'informazione clinica completa ed essenziale di pazienti con diagnosi nota, archiviata ed aggiornata in maniera sistematica ad ogni episodio di cura della storia clinica del paziente, allo scopo di identificare il contributo di determinanti ambientali al

fenotipo clinico, l'andamento temporale e la storia naturale della malattia, l'intervallo di confidenza per le prescrizioni del percorso assistenziale, la risposta al trattamento, il follow-up e la prognosi per ogni paziente affetto.

Il modulo CCI deve possedere caratteristiche trasversali rispetto alle cartelle informatizzate ospedaliere, dedicate ad un singolo episodio di cura e strutturarsi sul modello emergente dei record sanitari individuali informatizzati, quali ad esempio in Italia, il Fascicolo Sanitario Elettronico individuale.

Il valore aggiunto dello strumento è l'integrazione delle potenzialità offerte dalla struttura del database relazionale di dati fenotipi e genotipici con le applicazioni del modulo CCI nella gestione pratica del percorso assistenziale e di cura dei malati rari, con caratteristiche vantaggiose di facile gestione, sostenibilità futura e possibilità di ulteriore implementazione.

2.3 Implementazione dei prodotti in diversi contesti reali di assistenza clinica d'eccellenza

Obiettivo ulteriore del progetto di dottorato è il miglioramento dell'accessibilità e dell'utilizzo in condivisione dell'informazione clinica, genetica ed assistenziale derivante dall'utilizzo dello strumento di cui agli obiettivi 2.1 e 2.2., attraverso l'inserimento in rete ad accesso controllato intranet del modulo CCI. I collegamenti in rete sono strutturati anche in base alla prescrizione assistenziale del paziente con malattia rara, con attivazione di risorse per presa in carico sanitaria e sociale, a partire dalle strutture ospedaliere, fino ai presidi a livello territoriale e domiciliare, in diversi contesti d'assistenza clinica d'eccellenza, quali la rete dei Centri Accreditati aderenti al progetto RARER, riconosciuti come esperti d'eccellenza e riferimento in Regione Emilia Romagna per diagnosi e cura di specifiche patologie rare.

Lo strumento descrittivo dell'informazione clinica fenotipica e genotipica si sviluppa, a partire dal modello di registro già in uso in regione Veneto per le malattie rare, attraverso creazione ed integrazione di 8 registri di patologia o gruppi di patologie rare correlate presenti nel registro della regione Veneto, d'interesse specifico nell'ambito del programma di ricerca Regione-Università RARER.

Nell'ambito del progetto ministeriale UnRareNet sono invece auspicabili l'implementazione del modulo CCI integrato al database relazionale, con ampliamento

del numero di malattie rare considerate e l'incremento della popolazione di pazienti del database relazionale con coinvolgimento di CA e risorse operanti nel progetto e la disponibilità di casistica sufficientemente ampia per creazione di un sistema esperto di analisi dei dati in grado di determinare una corrispondenza univoca fenotipo-genotipica per le singole patologie rare nei pazienti con diagnosi nota e fornire una preselezione di casi, con attribuzione di un profilo di probabilità diagnostica a priori a ciascun caso e validazione della diagnosi attraverso i dati di pazienti con diagnosi nota.

Il sistema dovrà permettere la raccolta di nuovi casi senza diagnosi, selezionati dalle reti per malattie rare, e dell'informazione clinica fenotipo-genotipica correlata derivante anche dall'applicazione di moderne tecniche di sequenziamento genomico, e lo studio della correlazione tra nuova variante genomica e differenti quadri fenotipici, per formulare nuove diagnosi di patologia nota o identificare eventualmente nuove entità nosologiche.

3 MATERIALI E METODI

3.1 Il disegno sperimentale

Il progetto di ricerca ha previsto un disegno sperimentale strutturato in 3 diverse fasi:

- 1) identificazione e condivisione, all'interno del protocollo comune RARER, in collaborazione con CA per malattie rare della regione Emilia Romagna ed in relazione a 8 patologie o aree di patologie rare correlate presenti nel registro della regione Veneto, delle risorse utilizzabili e della metodologia idonea per:
 - a) implementazione della banca dati clinica, nella sezione relativa alla raccolta di segni e sintomi, secondo una logica gerarchica con suddivisione per organi ed apparati,
 - b) strutturazione di un modulo genotipo, una banca dati genomica per classificazione delle entità nosologiche secondo una logica integrata, che consenta la ridefinizione delle varianti genomiche in relazione al fenotipo clinico, all'efficacia dei trattamenti ed alla prognosi;
- 2) realizzazione e condivisione in rete del modulo CCI come prodotto sostenibile ed utile alla gestione clinica dei malati rari, contenente l'informazione clinica descritta all'interno di un database relazionale a struttura gerarchica ad elevato livello di granularità o aggregazione;
- 3) implementazione sistema d'acquisizione e di codifica dell'informazione clinica e del modulo CCI con:
 - l'ampliamento della popolazione del database, con inserimento dei casi con diagnosi nota o sospetta da parte dei CA aderenti al progetto UnRareNet;
 - l'analisi dei dati, con identificazione ove possibile dell'informazione clinica complessa sulla correlazione tra quadri fenotipici e genotipo;
 - l'istruzione di un sistema esperto con capacità analitiche e di auto-apprendimento in grado di gestire procedure complesse di elaborazione dati.

3.2 Le variabili e la logica classificatoria

Il progetto di ricerca ha previsto la realizzazione di strumenti descrittivi e di prodotti utili in ambito clinico, considerando le variabili seguenti:

- i pazienti affetti, o probabilmente affetti, da alcune malattie rare, o gruppi di malattie rare correlate, comprese nel Registro Regionale per le Malattie Rare del Veneto e di specifico interesse nell'ambito dei due più ampi progetti di cui a punto 3.1;
- i dati clinici, ovvero le caratteristiche fenotipiche e genotipiche, della popolazione di pazienti considerata.

Le variabili pazienti sono descritte all'interno di un database relazionale secondo differenti caratteristiche, raccolte attraverso il modulo CCI, rispettando una logica classificatoria a struttura nodulare suddivisa in differenti sottomoduli specifici, alcuni dei quali contenenti a loro volta un numero consistente di entità principali o secondarie tra loro connesse da altrettanto numerose relazioni.

Le variabili dati sono descritte nel database relazionale e classificate secondo i seguenti differenti livelli logici:

- le entità principali nosologiche (diagnosi certa o sospetta) sono descritte secondo una logica classificatoria con struttura multigerarchica ad albero, in grado di definire un crescente livello di granularità, all'interno di cui ogni entità nosologica è autoesplicativa, unica e distinta; ogni percorso logico (clinico, fisiopatologico o anatomico, terapeutico, eziologico, genomico etc.), rappresenta una ramificazione principale o secondaria della struttura ad albero ed ogni nodo rappresenta invece un'entità nosologica, che mapperà a differenti livelli della ramificazione (clinico a livelli più terminali o alti ad esempio e genomico a livelli più apicali o profondi) e potrà pertanto essere rappresentata all'interno di più ramificazioni, a seconda dei differenti percorsi logici sottostanti (parentela multipla), restando comunque sempre rintracciabile all'interno della classificazione stessa. L'altra caratteristica cardine del database relazionale è infatti la rintracciabilità delle singole entità nosologiche, che rappresentano il concetto centrale dell'intera logica del database; tale

caratteristica viene costantemente garantita da un codice classificativo specifico ed univoco, che si riferisce sempre alla medesima entità e rimane costante nel tempo;

- le entità principali e secondarie descriventi le caratteristiche fenotipiche dei pazienti risultano organizzate analiticamente, secondo una logica classificatoria gerarchica che rispetta la suddivisione anatomica e funzionale dell'organismo umano: le entità principali segni sono descritte secondo le variabili anatomo-funzionali sistema, alterazione, segno; le entità principali sintomi sono descritte secondo le variabili anatomo-funzionali sistema, sintomo; l'entità secondaria attributo di localizzazione prevede una serie molto numerosa di valori, corrispondenti alla localizzazione anatomica umana in organi ed apparati, con il maggior grado di completezza possibile e fino al massimo dettaglio (istologico-citologico);
- anche le entità principali esiti delle indagini diagnostiche risultano organizzate secondo una struttura ad albero con gruppi e sottogruppi (foglie) d'indagini specifiche.

3.3 Entità e relazioni

Il database relazionale contiene, organizzate per ciascun paziente secondo il modulo CCI ed i relativi sotto-moduli le seguenti classi di oggetti:

- entità principali: entità nosologica (sospetta o definitiva), sintomi, segni, comorbidità, menomazioni o danni funzionali, epicrisi, indagini diagnostiche (accertamenti ematologici, bio-chimici, metabolici, diagnostica per immagini, istologia, etc.), genotipo;
- entità secondarie: attributi principali comuni quali ad esempio contenuti, data di rilevazione, evoluzione clinica, severità, attributi principali specifici quali localizzazione, grandezza e altre caratteristiche quali numero, consistenza, colore, CA di riferimento etc.;
- relazioni multiple tra entità principali e secondarie, caratterizzate da un numero molto elevato di legami complessi uno a molti e molti a molti di ciascuna entità principale o secondaria con ciascuna altra.

3.4 Le fonti

La realizzazione dello strumento di descrizione analitica e precodificata dell'insieme di diversi fenotipi e genotipi di una popolazione di pazienti affetti o probabilmente affetti da malattie rare e del modulo CCI è partita dal modello unico già in uso del Registro Regionale per le Malattie Rare del Veneto.

La creazione e la manutenzione dell'infrastruttura informatica per la raccolta dati, dell'organizzazione dei sistemi di classificazione dell'informazione ottenuta e di analisi dei dati, il coordinamento delle attività dei Centri di riferimento coinvolti e la supervisione del lavoro di sviluppo, codifica, analisi logica, selezione e trattamento dei contenuti del database si sono realizzati attraverso le competenze specifiche dei clinici e degli operatori del Centro Coordinamento Regionale per le Malattie Rare della Regione Veneto.

La classificazione delle entità nosologiche ha rispettato le caratteristiche dei sistemi classificativi nosologici riconosciuti a livello internazionale, o in alternativa ha integrato una logica gerarchica condivisa con altri sistemi classificativi internazionali.

I sistemi classificativi considerati sono:

- la classificazione ICD, gerarchica con caratteristiche di unidimensionalità e con livello di granularità limitato al profilo di aggregazione dei gruppi di malattie correlate;
- la classificazione OMIM, gerarchica con caratteristiche di unidimensionalità e con livello di granularità non dipendente dal profilo di aggregazione dei gruppi di malattie correlate;
- la classificazione Orphanet, con caratteristiche di multidimensionalità, che garantisce l'associazione di ciascuna entità con un codice univoco (numero Orpha) anche in caso di parentela multipla.

Per la codifica e descrizione delle entità principali appartenenti ai moduli sezione controlli e diagnostica, ovvero quelle descrittive le caratteristiche del fenotipo e gli esiti delle indagini diagnostiche, ha sfruttato le risorse di database clinici Orphanet, accessibili al sito www.orpha.net e OMIM, accessibili al sito www.omim.org.

La descrizione delle entità descrittive caratteristiche fenotipiche e gli esiti delle indagini diagnostiche, relativamente al gruppo di patologie rare considerate nel progetto, è stata effettuata attraverso la revisione della principale letteratura internazionale, con particolare riferimento agli studi clinici e di popolazione con pazienti affetti da malattia rara o a studi epidemiologici con misure di frequenza di segni e sintomi in diversi profili fenotipi; la consultazione della letteratura scientifica è stata condotta tramite le risorse NCBI PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>).

L'implementazione delle banche dati cliniche, con perfezionamento delle metodologie descrittive delle caratteristiche del fenotipo e completamento dei moduli relativi alle indagini è stata realizzata attraverso indispensabili competenze specialistiche, condivise dai clinici operanti nei CA per malattia rara, coinvolti nei progetti RARER e UnRARENet.:

- Azienda Ospedaliero-Universitaria-Policlinico di Modena, centro delle Malattie Eredometaboliche del fegato
- Azienda Ospedaliero-Universitaria-Policlinico di Modena, Unità di Genetica Medica
- Azienda Ospedaliero Universitaria di Ferrara
- Istituto Ortopedico Rizzoli di Bologna
- Dipartimento di Scienze Biomediche dell'Università di Modena e Reggio Emilia, Unità Operativa di Genetica Medica
- Dipartimento di Scienze Biomediche dell'Università di Modena e Reggio Emilia, Unità Operativa di Genetica dei Disordini dei Lipidi
- Azienda Ospedaliero-Universitaria di Parma, Unità di Genetica Medica
- Azienda Ospedaliero-Universitaria di Parma, Unità di Genetica Molecolare
- Ospedale Pediatrico Bambino Gesù
- Azienda Ospedaliera Federico II di Napoli, Dipartimento di Pediatria
- Istituto Neurologico Carlo Besta di Milano Unità Operativa Complessa di Neurogenetica
- Istituto Dermatologico dell'Immacolata- IDI- IRCCS Roma.

I CA inoltre sono responsabili, attraverso accesso regolamentato alla rete informativa dedicata, dell'inserimento dei dati paziente secondo il modulo unico della CCI.

3.5 I contenuti ed i prodotti

Il database relazionale contiene dati clinici analitici e precodificati per ciascun paziente del campione di popolazione con diagnosi nota o sospetta di malattia rara, nell'ambito delle patologie di cui al punto 3.2., raccolti attraverso l'applicazione clinica modulo CCI e dei seguenti sotto-moduli specifici, descrittivi delle caratteristiche fenotipiche e genotipiche del paziente, gli esiti d'indagini diagnostiche ed i trattamenti pregressi rilevanti: anagrafica (dati paziente e diagnosi certa o sospetta), albero entità nosologie, piano terapeutico assistenziale, sezione controlli (fenotipo e trattamenti pregressi, genotipo), diagnostica.

La tecnologia informativa che supporta lo strumento è Piattaforma Cloud privata basata su infrastruttura hardware a tre livelli, completamente ridondante e clusterizzata.

Il database adottato è di tipo Oracle, con possibilità di gestione multipla di host Vmware e VSphere, si sviluppa all'interno del data center della Regione Veneto ed è collocato su differenti network proprietari SPC o GARR per ogni funzionalità: web server, application layer e database server. Il sistema è provvisto di backup automatici, che garantiscono il continuo funzionamento del sistema e permettono l'archiviazione ottica dei dati in caso di disaster recovery.

Il prodotto finale del modulo CCI è un'interfaccia utente di tipo JAVA, che rispetta la struttura nodulare in sotto-moduli di cui al punto 3.2, contenenti le entità principali e secondarie per identificazione del paziente, della diagnosi certa o del sospetto diagnostico, e per descrizione delle caratteristiche analitiche del profilo fenotipico, dei trattamenti pregressi rilevanti ed dei test genetici.

L'elaborazione dati contenuti nel database ha prodotto:

- una tabella formato Excel con distribuzione per righe dei pazienti e per colonne di una serie di variabili paziente specifiche;

- diverse tabelle formato Excel con distribuzione per righe di variabili entità principali e secondarie e per colonne di variabili entità principali e secondarie, mostrandone le specifiche relazioni.

La raccolta delle misure teoriche utili all'istruzione di un sistema esperto ha prodotto:

- diverse tabelle formato Excel, relative alle entità nosologiche del gruppo di malattie metaboliche considerate, con le variabili segno clinico-esito d'indagine diagnostica distribuite per righe e le variabili frequenza percentuale e fonte bibliografica di riferimento distribuite per colonne. Le frequenze teoriche dei principali segni clinici indicati in letteratura sono state quindi inserite nel sistema esperto per configurare casi simulati per alcuni quadri sindromici d'interesse;
- alcune tabelle Excel con i casi simulati distribuiti per righe e le variabili segno distribuite per colonne.

3.6 L'implementazione

La prima fase di utilizzo e sviluppo del strumento descritto, con inserimento dati di un non elevato numero di pazienti, ha permesso di verificare la tenuta del sistema in condizioni di fenotipi simili e dissimili.

L'implementazione del sistema avviene attraverso il continuo aggiornamento del modulo CCI, con periodiche modifiche basate su esigenze operative via via emergenti, riguardanti i sempre maggiori dettagli descrittivi delle caratteristiche fenotipiche ed i work-up assistenziali in atto per per i singoli pazienti arruolati nel database.

L'implementazione del sistema d'acquisizione e di codifica dell'informazione clinica e del modulo CCI può avvenire attraverso:

- l'ampliamento della popolazione del database, con inserimento dei casi con diagnosi nota o sospetta da parte dei CA aderenti al progetto UnRareNet;
- l'analisi dei dati, con identificazione ove possibile dell'informazione clinica complessa sulla correlazione tra quadri fenotipici e genotipo;

- l'istruzione di un sistema esperto con capacità analitiche e di auto-apprendimento i grado di gestire procedure complesse di elaborazione dati, che consentano di attribuire a priori di un profilo di probabilità diagnostica a ciascun caso selezionato e di validarne la diagnosi attraverso i dati di pazienti con diagnosi nota;

L'operatività del sistema esperto, nella prima fase di acquisizione pazienti è stata verificata per l'attribuzione di profili di probabilità diagnostiche a casi simulati tramite programma statistico, utilizzando misure di frequenza delle principali manifestazioni cliniche delle patologie rare selezionate, derivanti da una revisione completa della letteratura scientifica internazionale.

3.7 I metodi di analisi

Direttamente dall'interfaccia utente del modulo CCI è stata estratta un serie d'immagini a scopo illustrativo del prodotto finale, corrispondenti alle schermate principali dei differenti sottomoduli della CCI e ad alcune schermate secondarie che illustrano le modalità di compilazione ed i percorsi guidati per l'utente utili alla navigazione all'interno del sistema.

I dati relativi ai pazienti arruolati, derivanti dalla compilazione del modulo CCI da parte dei singoli operatori dei CA e raccolti nel database relazionale, sono stati processati ed estratti tramite il software Oracle Sequel Developer (SQL Developer), in modo da renderli automaticamente disponibili in formato Excel per successive elaborazioni.

I dati contenuti nelle tabelle formato Excel sono stati analizzati per identificazione di misure statistiche descrittive dei contenuti del database relazionale, entità principali e secondarie e relazioni e di caratteristiche utili alla definizione del campione di pazienti considerato.

I dati relativi alla popolazione del campione considerato sono stati inoltre sottoposti a semplici procedure elaborative per verificare l'effettiva usabilità e tenuta del sistema nell'arruolamento pazienti da parte dei singoli CA operanti ed una serie d'informazioni utili chiarire le potenzialità descrittive del sistema prodotto, in particolare in relazione alle numerose malattie rare di cui risultano affetti o

probabilmente affetti i pazienti del campione considerato e, più in generale, la validità del modello descrittivo e di raccolta dell'informazione clinica con potenziale riferimento a tutte le patologie umane note comuni o rare.

Nella prima fase d'istruzione del sistema esperto le misure di frequenza dei principali segni clinici e dei reperti diagnostici, relativi a selezionati gruppi di malattie rare e derivanti dalla letteratura internazionale sono state inserite nel sistema e sottoposte ad analisi tramite software integrato Statistical Analysis System (SAS), per descrizione di casi clinici virtuali ed attribuzione di profili di probabilità diagnostiche a priori da validare attraverso i dati dei casi con diagnosi nota.

4 RISULTATI

4.1 Contenuti del modulo CCI

Il modulo CCI è stato suddiviso nei seguenti sotto moduli:

- Anagrafica (dati paziente e diagnosi certa o sospetta);
- Albero entità nosologie;
- Piano terapeutico assistenziale e trattamenti pregressi;
- Sezione controlli (fenotipo e genotipo);
- Diagnostica.

Di seguito sono riportate 10 figure tratte dall'interfaccia utente del modulo CCI, che consente l'inserimento dei dati del paziente attraverso navigazione guidata per schermate principali e secondarie, che rispettano la suddivisione logica in sotto moduli della struttura CCI.

The screenshot shows the 'Registro Malattie Rare' interface. The main title is 'Registro Malattie Rare' and the sub-section is 'Malattie Rare'. The user is logged in as 'Utente connesso: mr_centro_pedpa'. The form is titled 'Dati del Paziente' and contains the following fields:

- Identificato - DATI PROVVISORI**
 - Cognome: ROSSATO
 - Nome: GABRIELLA
 - Detto: []
 - Data di nascita: 06/07/1976
 - Sesso: Maschile Femminile
 - Comune di nascita: PADOVA
 - Prov.: PD
 - Stato estero di nascita: []
 - Codice fiscale / STP: []
 - Codice sanitario: []
- Indirizzo di residenza**
 - Indirizzo: VIA GIGLIO 85
 - Cap: []
 - Comune di residenza: PADOVA
 - Prov.: PD
 - Regione di residenza: Veneto
 - Azienda ULSS di residenza: PADOVA
 - Stato estero di residenza: []
- Indirizzo di domicilio**
 - Domicilio diverso dalla residenza: Sì No
 - Indirizzo: []
 - Cap: []
 - Comune di domicilio: []
 - Prov.: []
 - Regione di domicilio: []
 - Azienda ULSS di domicilio: []
 - Stato estero di domicilio: []
 - Assistenza fornita da: Azienda ULSS di domicilio Azienda ULSS di residenza
 - Spedire esenzione a domicilio: Sì No

Figura 1: Schermata principale del modulo anagrafica parte a

Lookup Table -- Finestra di dialogo pagina Web

http://malattierare-test.test.pd/jsp/SWGeneralizzato/lookup.jsp?dato=_malattia&functionName=setMalattii

<< < > >> Esci Pag. 1

	Mal. rif.	cod. esenz	cod. ICD9CM
S SINDROME DI ALPORT	SINDROME DI ALPORT	RN1360	759.89
S SINDROME DI ALPORT AUTOSOMICA DOMINANTE	SINDROME DI ALPORT	RN1360	759.89
S SINDROME DI ALPORT AUTOSOMICA RECESSIVA	SINDROME DI ALPORT	RN1360	759.89
S SINDROME DI ALPORT X LINKED	SINDROME DI ALPORT	RN1360	759.89

Figura 2: Schermata principale del modulo "Albero delle malattie"

Registro Malattie Rare

Malattie Rare

Nuovo Certificato

Cerca Pazienti

Cerca Paziente con PIN

Ricerca nei Piani Terapeutici

Estrazione Dati

Gestione Somministrazioni Orfani

Gestione Somministrazioni Orfani

Elaborazione Guidata SAS

Segnalazione Malattie non comprese

Elenco Malattie non comprese

Cambio Password

Manuale

Piano Terapeutico Personalizzato

Utente connesso: mr_centro_pedp

Prescrizione rilasciata al paziente il 21/11/2011 da Dr.ssa mr_centro_pedp e valida fino al 21/11/2012

1. Trattamenti extra-LEA prescrivibili gratuitamente in base a specifici protocolli regionali

Terapia	Nome commerciale	Posologia	Off-Label	Non-Sost	Scheda
PIRIDOSSINA	BENADON*10CPR GASTRORES 300MG 1 cp die		<input type="checkbox"/>	<input type="checkbox"/>	

Farmacia che eroga: S.S. Farmacia Territoriale - Ospedale S. Antonio

2. Farmaci e parafarmaci

Principio Attivo	Nome Commerciale	Posologia	Nota-AIFA	Off-Label	Non-Sost	Scheda
ACIDO FUSIDICO	FUCIDINE 250 MG 10 CPS	1 cpr die	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	mostra
	COLECALCIUM SCIR 150ML	mmm	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
BETAINA	CYSTDANE*FOS 1FL. 180G+3CUCCH	sss	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	mostra

Farmacia che eroga: S.C.I. Farmacia Ospedaliera - Azienda Ospedaliera di Padova

3. Richiesta di trattamento/i extra-LEA essenziali ed insostituibili da autorizzare in regime di esenzione dal costo (per i residenti nelle ULSS della Regione Veneto in base alla nota prof. 305255/50070202 dell'11/06/2000)

Principio Attivo	Nome Commerciale	Posologia	Off-Label	Non-Sost	Galen.	Relaz.	Scheda
ACIDO ASCORBICO	CEBION 500*20CPR MAST ARANCIA	1 cpr die	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	mostra

Farmacia che eroga: S.C.I. Farmacia Ospedaliera - Azienda Ospedaliera di Padova

4. Preparati galenici

Principio Attivo	Nome Commerciale	Posologia	Off-Label	Non-Sost	Scheda
RIBOFLAVINA (galenico)		bs x mg	<input type="checkbox"/>	<input type="checkbox"/>	mostra

Farmacia che eroga: S.C.I. Farmacia Ospedaliera - Azienda Ospedaliera di Padova

6. Altri trattamenti (riabilitativi, chirurgici, etc.)

Trattamento	Data
fisioterapia	12/12/2011

7. Note del prescrivente

vvv

Elenco PTP Rinnova Modifica Interrompi Nuovo piano Ristampa

Figura 3: Schermata principale del modulo "Piano Terapeutico Assistenziale"

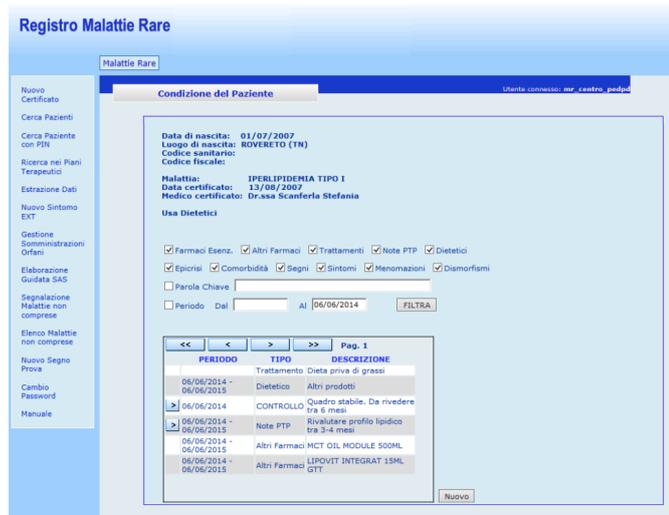


Figura 4: Schermata principale del modulo "Controlli"; possibilità di filtrare le caratteristiche fenotipiche

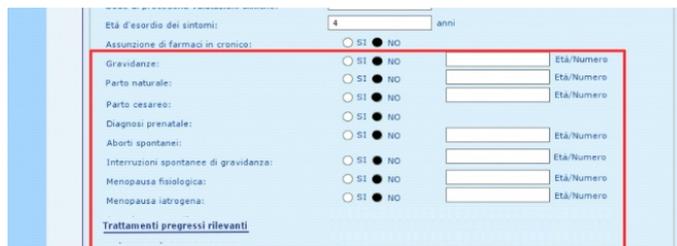


Figura 5: Modulo Controlli; apertura schermata secondaria di compilazione di altre caratteristiche fenotipiche e dei trattamenti progressi rilevanti

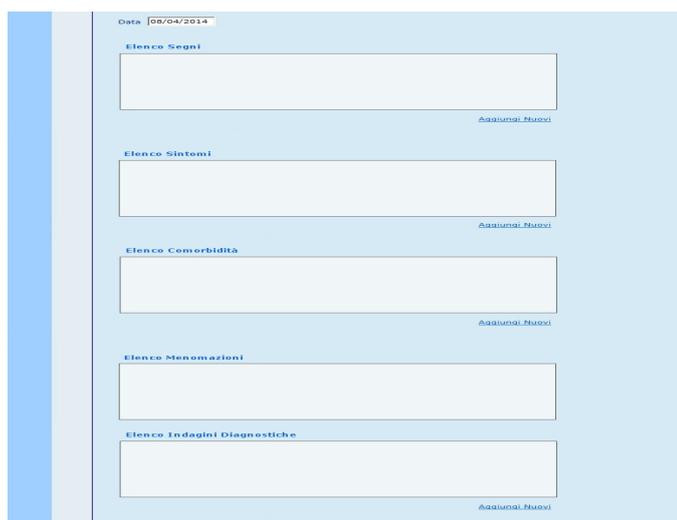


Figura 6: Modulo Controlli; sotto sezioni con entità principali, descriventi il fenotipo

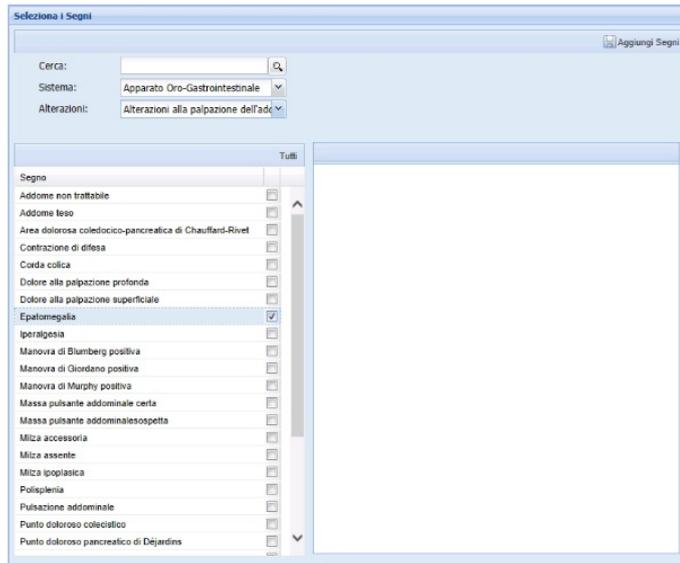


Figura 7: Modulo Controlli; sotto sezione con entità principali segni

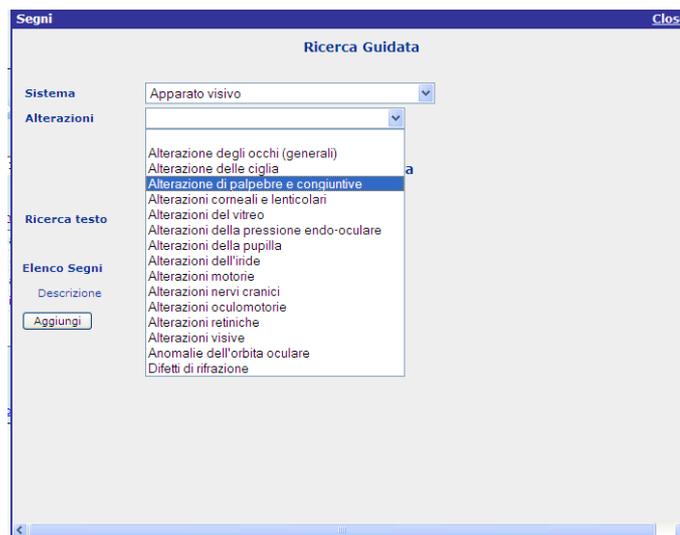


Figura 8: Modulo Controlli; schermata secondaria di ricerca segni nel thesaurus a logica gerarchica

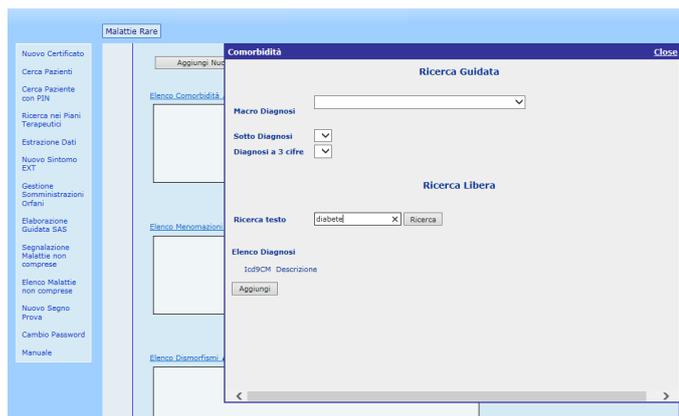


Figura 9: Modulo Controlli; sotto sezione comorbidità



Figura 10: Modulo Controlli; sotto sezione genotipo

4.2 Contenuti del database relazionale

4.2.1 Entità e relazioni

Le tabelle Excel che riportano i dati contenuti nel database non sono mostrate integralmente per ragioni di spazio; di seguito sono descritti il nome ed il contenuto delle tabelle con la distribuzione delle variabili per righe e colonne; è inoltre riportato il numero totale di righe presenti.

“ElencoSegni” riporta entità principali segni distribuiti per righe ed i dati di ciascuna entità distribuiti per colonna secondo le variabili: ID progressivo e nome di sistema, alterazione e segno. Il database contiene 5045 entità principali segni.

“ElencoSintomi” riporta entità principali sintomi distribuiti per righe ed i dati di ciascuna entità distribuiti per colonna secondo le variabili: ID progressivo e nome di sistema, e sintomo. Il database contiene 295 entità principali segni.

“ElencoComorbidità” riporta le entità principali comorbidità distribuite per righe ed i dati di ciascuna entità distribuite per colonna secondo le variabili: ID progressivo e nome, del macrogruppo e del microgruppo di appartenenza nel sistema classificativo ICD-9 CM. Il database contiene 13861 entità secondarie comorbidità.

“ElencoMenomazioni” riporta le entità principali menomazioni distribuite per righe ed i dati di ciascuna entità distribuite per colonna secondo le variabili: ID progressivo e nome del macrogruppo e del microgruppo di appartenenza nel sistema classificativo ICD-9 CM, descrizione e codice della menomazione. Il database contiene 278 entità secondarie menomazioni.

“ElencoIndagini” riporta le entità principali indagini diagnostiche distribuiti per righe ed i dati di ciascuna entità distribuite per colonna secondo le variabili: ID progressivo e nome di macrogruppo, microgruppo e indagine diagnostica. Il database contiene 295 entità principali segni.

“ElencoAttributi” riporta le entità secondarie attributi distribuite per righe ed i dati di ciascun attributo distribuiti per colonna secondo le variabili: ID progressivo, nome, tipo e descrizione della voce di compilazione e della voce di riepilogo dell’attributo. Il database contiene 903 entità secondarie attributi.

“ElencoLocalizzazioni” riporta le entità secondarie attributi di localizzazione distribuite per righe ed i dati di ciascuna entità distribuite per colonna secondo le variabili: ID progressivo e descrizione della localizzazione, livello numerico corrispondente nella struttura ad albero delle localizzazioni, presenza o assenza di lateralizzazione. Il database contiene 18201 entità secondarie attributo di localizzazione.

“ElencoSegniLocalizzazioni” mostra la relazione tra le entità segni e le entità attributi di localizzazione, riportando le entità principali segni distribuite per righe ed i dati di ciascuna entità distribuiti per colonna secondo le variabili: ID e nome di sistema, alterazione, segno, presenza e descrizione o assenza della localizzazione, livello numerico corrispondente nella struttura ad albero delle localizzazioni, presenza o assenza di lateralizzazione. Il database contiene 5006 relazioni tra le entità principali segni e le entità secondarie attributo di localizzazione.

“ElencoSegniSistemiAlterazioniAttributi” mostra la relazione tra le entità segni, sistemi, alterazioni e le entità attributi, riportando le entità principali segni distribuite

per righe ed i dati di ciascuna entità distribuiti per colonna secondo le variabili: ID e nome di sistema, alterazione e segno, ID e nome dell'attributo. Il database contiene 25789 relazioni tra le entità principali segni e le entità secondarie attributi.

“ElencoSintomiLocalizzazioni” mostra la relazione tra le entità sintomi e le entità attributi di localizzazione, riportando le entità principali segni distribuite per righe ed i dati di ciascuna entità distribuiti per colonna secondo le variabili: ID e nome di sistema, sintomo, presenza e descrizione o assenza della localizzazione, livello numerico corrispondente nella struttura ad albero delle localizzazioni, presenza o assenza di lateralizzazione. Il database contiene 299 relazioni tra le entità principali sintomi e le entità secondarie attributo di localizzazione.

“ElencoSintomiSistemiAttributi” mostra la relazione tra le entità sintomi, alterazioni e le entità attributi, riportando le entità principali sintomi distribuite per righe ed i dati di ciascuna entità distribuiti per colonna secondo le variabili: ID e nome di sistema e sintomo, ID e nome dell'attributo. Il database contiene 2081 relazioni tra le entità principali sintomi e le entità secondarie attributi.

“ElencoIndaginiSegni” mostra la relazione tra le entità principali indagini diagnostiche e le entità principali segni, riportando le entità principali indagini diagnostiche distribuite per righe ed i dati di ciascuna entità distribuiti per colonna secondo le variabili: ID e nome di macrogruppo, microgruppo e indagine diagnostica, ID e nome del segno. Il database contiene 27525 relazioni tra le entità principali indagini diagnostiche e le entità principali segni.

Segni	4372
Sintomi	295
Comorbidità	13861
Menomazioni	278
Indagini diagnostiche	1275
Attributi	903
Localizzazioni	18201
Relazioni segni/localizzazioni	5006
Relazioni segni/attributi	25789
Relazioni sintomi/localizzazioni	299

Relazioni sintomi/attributi	2081
-----------------------------	------

Tabella 1: riepilogo numero totale di entità principali e secondarie e di relazioni contenute nel database relazionale

4.2.2 Pazienti

La tabella Excel che riporta i pazienti contenuti nel database non è mostrata integralmente per ragioni di spazio; di seguito sono descritti il nome ed il contenuto della tabella con la distribuzione delle variabili per righe e colonne; è inoltre riportato il numero totale di righe presenti.

“ElencoPazienti” riporta i pazienti distribuiti per righe ed i dati clinico-anamnestici distribuiti per colonna secondo le variabili: cognome, nome, sesso, data di nascita, telefono fisso, telefono cellulare, e-mail, ID del Comune di nascita, codice fiscale, indirizzo di residenza e ID Comune di residenza, codice Regione, ID e nome dell’Azienda Sanitaria Locale di residenza, ID e nome malattia rara, ID e nome sospetta malattia rara, ID e nome del centro, dell’azienda, del presidio e del dipartimento e del medico che certifica il paziente, data inserimento valori, ID identificativo utente per inserimento dati, data di certificazione, presenza e numero o assenza d’indagine genetica, segni, sintomi, comorbidità, menomazioni, presenza o assenza di epicrisi, presenza o assenza di consenso informato. Il database contiene al momento attuale 1186 pazienti.

Segni	Sintomi	Comorbidità	Menomazioni	Indagini	Indagine Genetica
2735	851	1030	231	1436	731

Tabella 2: Numero totale di entità principali riportate per ciascuno dei pazienti del campione

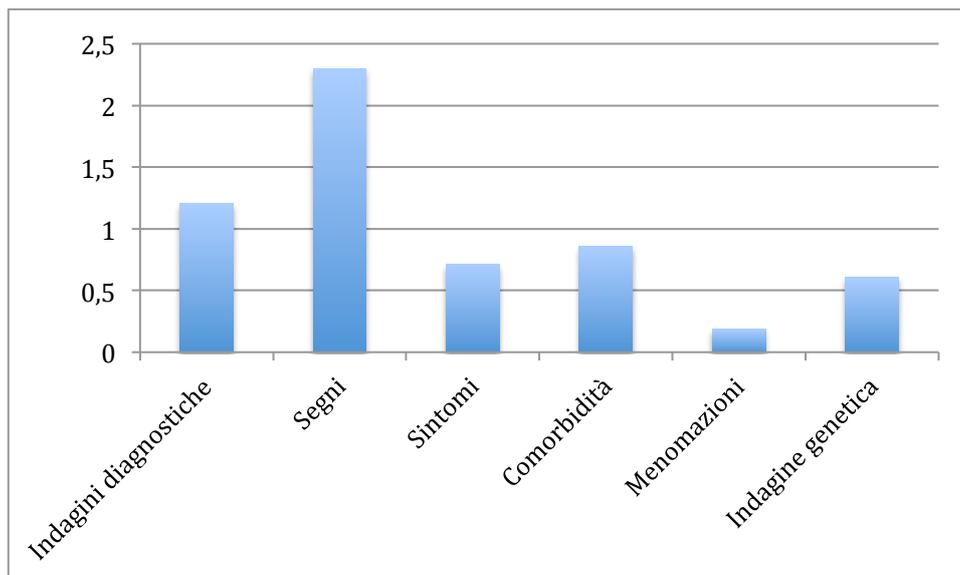


Figura 11: Distribuzione media per paziente delle entità principali

Dal grafico in Figura 11 notiamo come nel database siano contenuti mediamente più di 2 segni per paziente ed almeno una comorbidità per paziente.

Il trend di arruolamento dei pazienti ha mostrato un andamento generale in crescendo, passando da una popolazione di 19 pazienti nell'agosto 2014, a 649 pazienti nell'agosto dell'anno successivo, fino ai 1186 pazienti attuali.

La Tabella 3 e la Figura 12 seguenti mostrano nel dettaglio il trend di arruolamento pazienti da parte dei singoli CA.

ID	1219	1220	1221	1222	1223	1224	1225	1227	1228	1229	1230	1231	1232	totale
ago-14	19	0	0	0	0	0	0	0	0	0	0	0	0	19
set-14	61	15	0	10	0	0	1	5	0	0	0	0	0	92
ott-14	6	5	1	8	0	3	0	0	0	0	2	30	0	55
nov-14	0	0	0	0	1	0	0	6	0	0	0	2	0	9
dic-14	0	0	0	0	0	0	1	11	0	0	0	4	0	16
gen-15	0	0	0	0	0	5	1	2	1	0	1	0	0	10
feb-15	11	0	0	0	0	3	1	0	0	0	3	0	0	18
mar-15	46	0	0	10	0	0	0	12	0	0	0	0	0	68
apr-15	21	7	0	13	0	5	0	0	0	0	0	0	0	46
mag-15	9	43	0	20	0	0	26	0	0	0	0	0	1	99
giu-15	32	8	10	11	68	5	3	0	34	0	12	12	0	195
lug-15	0	0	5	1	0	0	1	8	0	0	0	6	0	21
ago-15	0	0	0	0	0	0	0	0	0	0	0	0	1	1
set-15	2	15	3	3	0	0	0	0	0	0	1	0	0	24

ott-15	0	12	1	4	0	0	26	0	0	0	5	0	341	389
nov-15	10	24	8	0	0	0	14	0	0	30	1	0	0	87
dic-15	0	18	5	0	0	0	0	5	0	0	2	0	0	30
gen-16	7	0	0	0	0	0	0	0	0	0	0	0	0	7
totale	224	147	33	80	69	21	74	49	35	30	27	54	343	1186

Tabella 3: Pazienti arruolati dai singoli CA nel corso del progetto

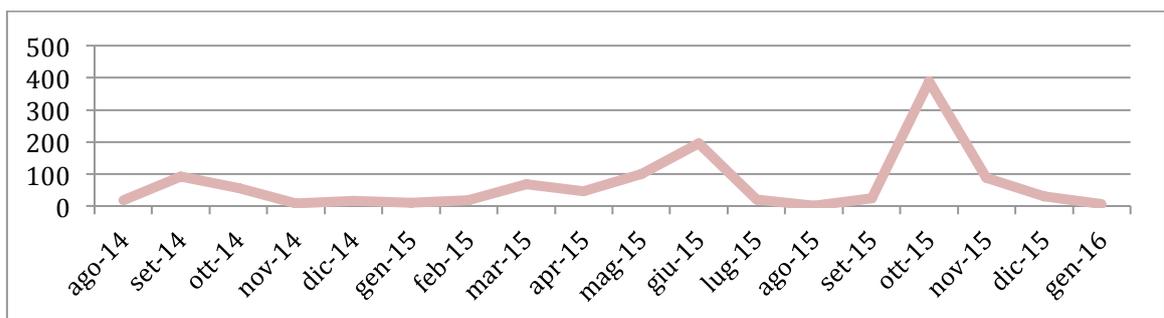


Figura 12: Andamento temporale dell'arruolamento pazienti nel corso del progetto

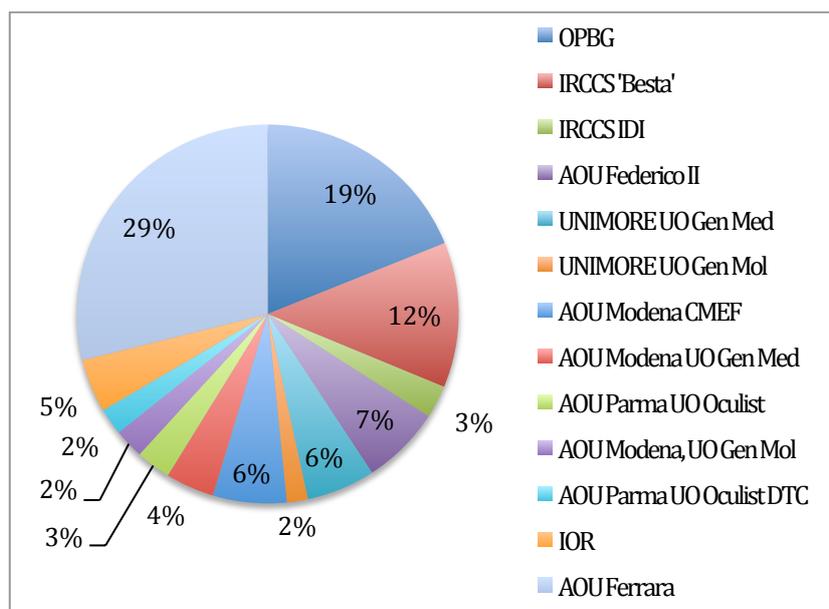


Figura 13: Distribuzione dei casi arruolati da ciascun CA

Tutti i pazienti risultano affetti o probabilmente affetti da una delle seguenti malattie rare (o entità comprese nei seguenti gruppi di malattie rare correlate) :

Accumulo marziale ereditario, Distrofia muscolare di Duchenne, Distrofia muscolare facio-scapolo-omerale (FSHD), Distrofia di Becker, Ipertrigliceridemie monogeniche, Ipoacusie ereditarie, Osteocondromatosi multipla, Neurofibromatosi,

Degenerazione retinica ereditaria, Malattie Rare dismorfologiche, Malattie Rare metaboliche, Malattie Rare Mitocondriali, le Malattie Rare Dermatologiche.

Il database contiene 117 diverse entità principali nosologiche, diagnosticate su un totale di 1065 pazienti, che non sono state riportate per ragioni di spazio. La malattia più rappresentata è la Distrofia muscolare di Duchenne, con 222 pazienti diagnosticati, seguita dalla Distrofia di Becker con 120 pazienti. Sono 53 le entità nosologiche meno frequenti, ciascuna con 1 solo paziente diagnosticato.

Distrofia Di Duchenne	222
Distrofia Di Becker	120
Sindrome Di Digeorge	87
Fshd	69
Esostosi Multiple Cartilaginee	53
Sindrome Melas	40
Emocromatosi Ereditaria	30
Sindrome Di Noonan	30
Sindrome Di Kearns Sayre	23
Sindrome Di Williams	23
Sordità Neurosensoriale Non Sindromica Autosomica Recessiva, Tipo Dfnb	22
Atrofia Ottica Autosomica Dominante	20
Sindrome Di Merrf	19
Retinite Pigmentosa	18
Sindrome Di Noonan E Sindromi Correlate	18
Neurofibromatosi Di Tipo 1	17
Atrofia Ottica Di Leber	15
Iperchilomicronemia Familiare	14
Sindrome Di Kabuki	11
Ittiosi Lamellare Recessiva	10

Tabella 4: Entità nosologiche maggiormente rappresentate per numero di pazienti diagnosticati, tra quelle contenute nel database relazionale

In generale il gruppo di malattie rare più rappresentato è quello delle Distrofie Muscolari Congenite, che rappresenta il 34% dei pazienti diagnosticati, seguito dal gruppo delle malattie rare dismorfologiche con il 19% dei pazienti, dal gruppo delle

malattie rare metaboliche con il 15% dei pazienti, dai gruppo delle malattie rare degli organi di senso (ipoacusie e degenerazioni retiniche ereditarie) e mitocondriali con il 12% dei pazienti, dal gruppo delle malattie rare scheletriche con il 5% dei pazienti e infine dal gruppo delle malattie rare dermatologiche con il 3% dei pazienti diagnosticati.

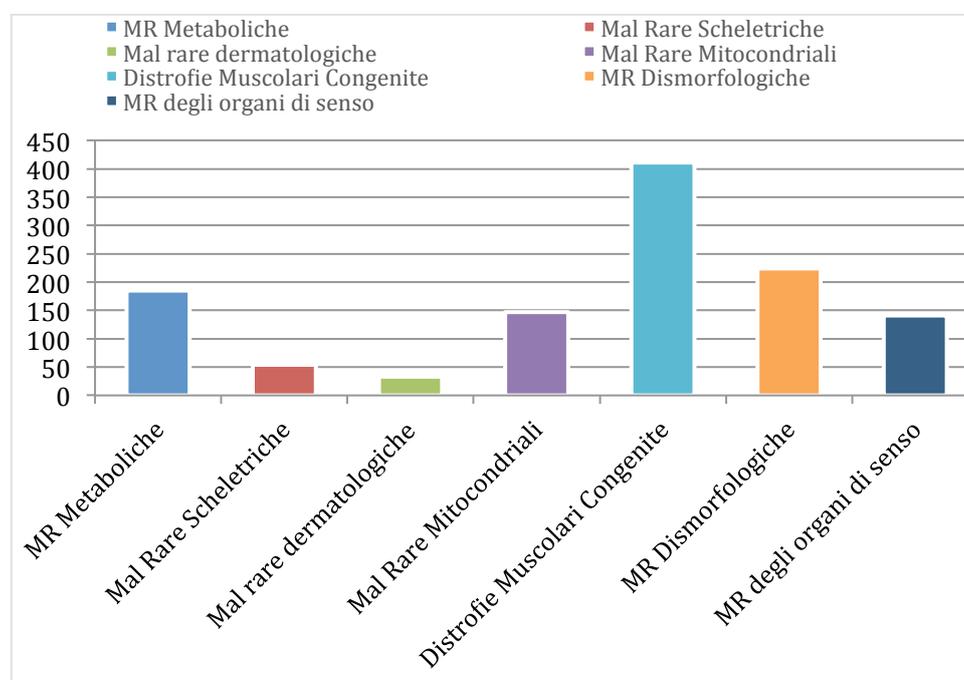


Figura 14: Numerosità dei pazienti per gruppi di malattie rare correlate

I pazienti del campione considerato presentano le seguenti caratteristiche: sesso (779 maschi e 407 femmine); età media 28,1 anni (range 0,5-85,5 anni) e distribuzione per età (235 pazienti con età compresa tra 0 e 10 anni, 255 pazienti con età compresa tra 11 e 20 anni, 154 pazienti con età compresa tra 21 e 30 anni, 127 pazienti con età compresa tra 31 e 40 anni, 122 pazienti con età compresa tra 41 e 50 anni, 83 pazienti con età compresa tra 51 e 60 anni, 69 pazienti con età compresa tra 61 e 70 anni, 20 pazienti con età compresa tra 61 e 80 anni, 5 pazienti con età compresa tra 81 e 90 anni).

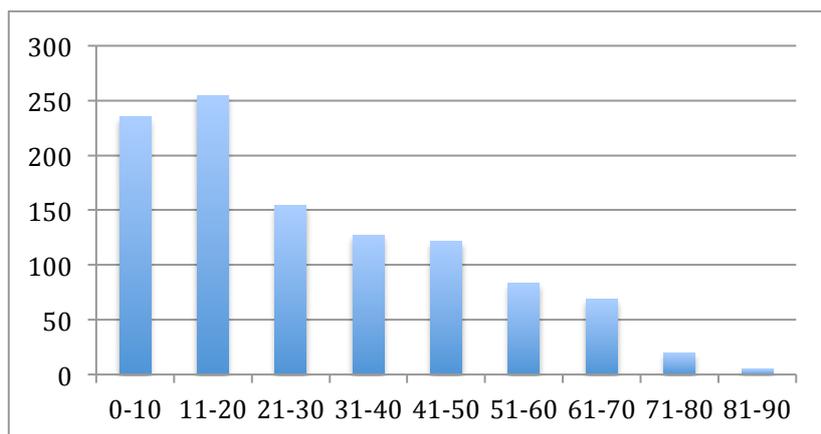


Figura 15: Distribuzione dei pazienti per fasce di età

Per 611/1186 pazienti è stata eseguita almeno un'indagine genetica: in 524/611 è stata eseguita una sola indagine genetica, in 87/611 sono state eseguite 2 o più indagini genetiche.

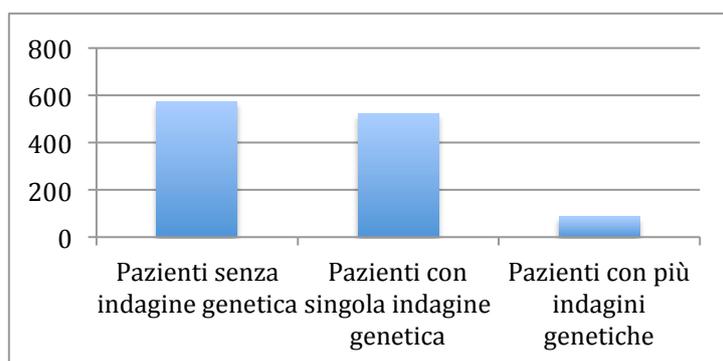


Figura 16: Numerosità dei pazienti senza e con una o più indagini genetiche

In totale per 611/1186 sono state eseguite complessivamente 870 indagini genetiche, con la seguente distribuzione:

Indagine genetica	Numero Pazienti
Analisi di Microsatelliti (VNTR, STR etc.)	1
Metodo colorimetrico	1
Quantitative Polymerase Chain Reaction (QFPCR)	1
SMP mediante arricchimento per cattura di reg. target	1
Metodo flurimetrico	2
Reverse Transcriptase Polymerase Chain Reaction RTPCR	8
Comparative Genomic Hybridation (CGH-Array)	11
altro	10
Sequenziamento Massivo Parallelo (SMP-NGS)	21
Restriction fragment length polymorphism (RFLP)	22

SMP di regioni selezionate mediante ampliconi	25
Fluorescent In-Situ Hybridation (FISH)	26
Cariotipo Metafasico	47
Real Time Polimerase Chain reaction- Real Time PCR	46
Denaturing High Performance Liquid Chromatography (DHPLC)	60
Digestione con enzimi di restrizione	63
Multiplex ligation-dependent Probe Amplification (MLPA)	66
Southern Blotting	91
Sequenziamento Sanger	289
Totale	791

Tabella 5: Numerosità pazienti per singole tipologie di indagine genetica specifica

Le 30 indagini citogenetiche (CGH-Array e FISH) hanno consentito l'identificazione di 30 alterazioni strutturali responsabili (delezioni) di 5 diverse malattie rare:

- delezioni del cromosoma 22q per Sindrome di DiGeorge;
- 23 delezioni del cromosoma 7q per la Sindrome di Williams;
- 1 delezione del cromosoma 8p per la Sindrome di Kabuki
- 1 delezione del cromosoma 3p, per la sindrome correlata;
- 1 delezione del cromosoma 16p, per la sindrome correlata.

Sono state eseguite 189 indagini di genetica molecolare, per le 17 complessive tecniche seguenti:

Sequenziamento massivo parallelo - SMP (Next Generation Sequencing)
Digestione con enzimi di restrizione
Southern Blotting
Sequenziamento Sanger
SMP mediante arricchimento per cattura di reg. specifiche(target)
SMP di regioni selezionate mediante ampliconi
Analisi di microsatelliti (es. VNTR, STR, etc.)
SMP dell'esoma mediante arricchimento per cattura
Multiplex Ligation-dependent Probe Amplification (MLPA)
Reverse Dot Blot
Reverse Transcriptase Polimerase Chain Reaction - RTPCR

Real-time Polimerase Chain Reaction- Real Time PCR
Quantitative Fluorescence Polimerase Chain Reaction - QFPCR
Restriction fragment length polymorphism (RFLP)
Denaturing High Performance Liquid Chromatography (DHPLC)

Tabella 6: Tecniche diagnostiche di indagine genetica molecolare

Le tecniche di genetica molecolare hanno consentito l'identificazione di 63 differenti geni, responsabili di altrettante patologie rare o gruppi di patologie rare correlate.

Le più frequenti patologie rare, identificate con le 189 indagini di genetica molecolare, sono le seguenti:

Gene	Malattia Rara	
PTPN11	Sindrome di Noonan	81
DNA mitocondriale	Malattie Rare Mitocondriali	75
NF1	Neurofibromatosi	50
GJB2	Ipoacusie ereditarie	49
EXT1	Esostosi Multipla	41
22q11.2	Sindrome di DiGeorge	39
HFE	Emocromatosi Ereditaria	28

Tabella 7: Numero di casi per patologia rara diagnosticati con indagini di genetica molecolare

121/1186 pazienti risultano senza diagnosi. In 1/121 pazienti senza diagnosi non è stato possibile formulare un sospetto diagnostico.

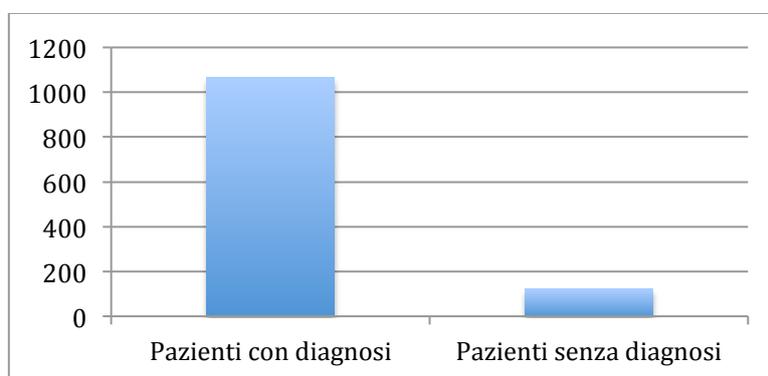


Figura 17: Distribuzione numerica pazienti per assenza o presenza di diagnosi nota

Per 73/121 pazienti senza diagnosi non risultano eseguite indagini genetiche, nei restanti 48 sono state eseguite una o più indagini genetiche e tra questi risulta compreso il paziente per cui non è stato possibile formulare un sospetto diagnostico (un'indagine genetica eseguita).

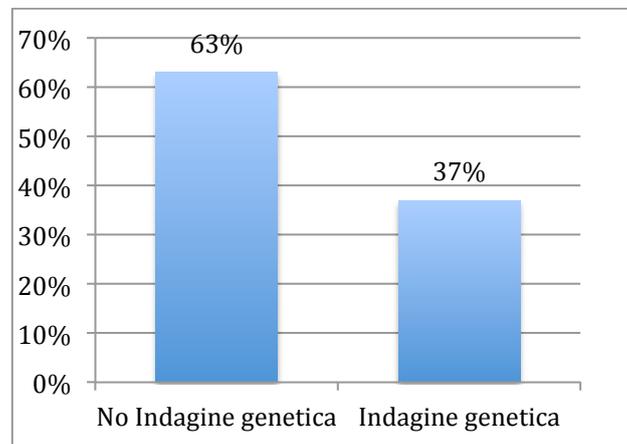


Figura 18: Distribuzione percentuale dei pazienti senza diagnosi per presenza o assenza di indagine genetica

Per i 1065/1186 con diagnosi nota, in 502/1065 casi non è stata eseguita alcuna indagine genetica.

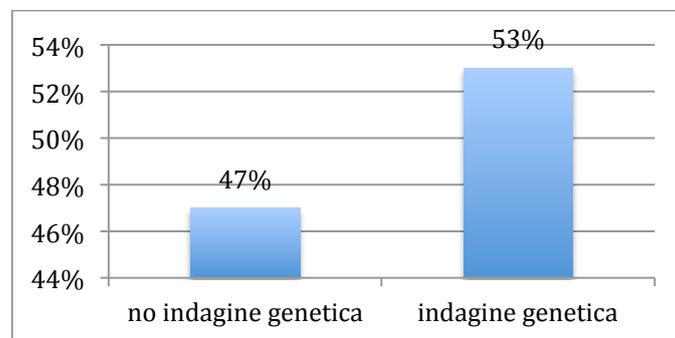


Figura 19: Distribuzione percentuale dei pazienti con diagnosi nota per presenza o assenza di indagine genetica

4.3 Il sistema esperto

Da un'analisi completa della letteratura internazionale sono state estratte misure teoriche di frequenza dei principali segni clinici e reperti diagnostici relativi alle seguenti malattie rare metaboliche o mitocondriali: Aciduria Argininsuccinica, Aciduria 2-Idrossiglutarica, Acidemia Isovalerica, Acidemie metilmaloniche, Acidemia metilmalonica e omocistinuria, Acidemia metilmalonica e omocistinuria, Tipo CBL C, Acidemie propioniche, Encefalopatia etilmalonica, Galattosemia, Gangliosidosi GM1

Infantile (Tipo 1), Intolleranza lisinurica alle proteine, Iperglicinemia non chetotica, Leucinosi, Malattia di CANAVAN, Malattia di FABRY, Malattia di GAUCHER, Malattia di KRABBE, Malattia di NIEMANN PICK Tipo C, Malattia di WOLMAN, Mucopolisaccaridosi Tipo II, Mucolipidosi Tipo III, Mucopolisaccaridosi Tipo IV, Mucopolisaccaridosi Tipo IVA, Mucopolisaccaridosi Tipo VI, Omocistinuria, Sindrome HHH, Sindrome DI SMITH-LEMLI-OPITZ Tipo I, Tirosinemia Tipo I, Glicogenosi Tipo IA, Glicogenosi Tipo IB, Glicogenosi Tipo II, Glicogenosi Tipo III, Sindrome di Leigh, sindrome MELAS, sindrome MERRF, sindrome Kearns Sayre, Atrofia ottica di Leber.

In Tabella 8 è riportato l'esempio della Acidemia Metilmalonica.

Segno clinico	Frequenza percentuale
Consanguineità	49%
Gemellarità	33%
Esordio in età neonatale	48%
Responsività alla cobalamina	19%
Non responsività alla cobalamina	67%
Crisi metabolica acuta	68%
Vomito ricorrente e difficoltà alimentari	13%
Deficit di crescita	9%
Altri segni	4%

Tabella 8: Frequenze percentuali teoriche per alcuni segni clinici principali della acidemia metilmalonica tratti da Horster et al. 2012.

Le tabelle Excel con i risultati dei casi simulati riportano, per ciascuna riga paziente, i valori 0 o 1, assegnati a caso, per indicare rispettivamente assenza o presenza di un determinato segno clinico, in base ai valori di frequenza teorica precedentemente inseriti per ciascun segno clinico. Il sistema ha quindi restituito i profili di probabilità attesi.

In tabella 9 è riportato l'esempio della Sindrome di Leigh.

Segno clinico	%
Ipotonia	74,6
Difficoltà alimentari	45,4
Distonia	44,6
Segno di Babinski	42,3
Ipertricosi	41

Tabella 9: Percentuali di frequenza teoriche per alcuni segni clinici principali della Sindrome di Leigh tratte da Safou et al. 2014

Le misure teoriche di frequenza estratte, inserite nel sistema ed elaborate tramite SAS, hanno restituito fino a 10.000 casi simulati di pazienti probabilmente affetti da una delle suddette patologie del gruppo metaboliche o mitocondriali, con corretta attribuzione di profili attesi di probabilità diagnostiche.

5 DISCUSSIONE E CONCLUSIONI

L'obiettivo generale del presente progetto di dottorato è stato la realizzazione uno strumento unico di raccolta dell'informazione clinica, ovvero di dati analitici descrittivi il fenotipo e il genotipo di pazienti con diagnosi nota o sospetta, all'interno di gruppi di malattie rare selezionate.

La sfida essenziale e critica è stata la creazione di una metodologia ed un sistema di raccolta dati nel rispetto di una logica basata sull'approccio clinico, che possa descrivere l'informazione clinica al maggior grado di completezza teoricamente possibile, nella specifica corrispondenza fenotipo-genotipica, ed in modo flessibile e non predefinito, ma aperto, completamente elaborabile e generalizzabile con utilizzo potenziale per tutte le malattie note di cui i soggetti arruolati potrebbero essere considerati affetti.

Lo strumento dovrebbe avere capacità adattative e potenzialità applicative anche rispetto ad uno scenario di ricerca, rispettare cioè la struttura di un database relazionale, in grado di gestire la raccolta aperta d'informazioni cliniche pre-codificate prontamente utilizzabili per elaborazioni complesse e rispondere all'ulteriore obiettivo di integrare l'informazione clinica e genetica (anche derivante dall'applicazione di tecniche di Next Generation Sequencing).

I risultati ottenuti nella fase di creazione ed implementazione del sistema di raccolta dell'informazione clinica, con la messa in atto del modulo CCI nel contesto del progetto RARER, hanno dimostrato la fattibilità del modello e l'usabilità dello stesso nello specifico ambito degli 8 registri di patologia inclusi nel progetto regione-università.

L'ampliamento dell'applicazione clinica nel contesto del progetto UnRareNet, ha prodotto risultati promettenti, rispetto alle esigenze di flessibilità del modello, ovvero capacità del sistema di acquisire nuove informazioni non ancorate alle conoscenze già note su specifiche entità nosologiche; a tal proposito il sistema è risultato in grado di gestire in modo completo informazioni cliniche riguardanti complessivamente 117 differenti malattie rare. Il modello ha altresì dimostrato potenzialità in ambito di ricerca

clinica, nello specifico i dati contenuti nel database relazionale hanno caratteristiche pre-codificate e sono risultati elaborabili ed inter-operabili in modo completo e rapido.

Si sono incontrate alcune difficoltà nelle prime fasi di applicazione clinica del modulo CCI, come dimostra l'andamento temporale dell'arruolamento pazienti, che ha subito una deflessione importante tra novembre 2014 e febbraio 2015, difficoltà verosimilmente attribuibili a ritardi e criticità nella creazione di moduli per l'inserimento di segni clinici o reperti diagnostici altamente specifici. A tal proposito sono stati fondamentali la condivisione di conoscenza ed il confronto con i clinici operanti nei singoli CA operanti nel progetto.

L'integrazione dell'informazione clinica e genetica appare fattibile e premettente, seppur nell'ambito del progetto UnRareNet sia ancora in fase assolutamente preliminare; nell'attuale popolazione di 1186 sono state eseguite prevalentemente indagini molecolari tradizionali, in qualche caso indagini di citogenetica sono state sufficienti, ma solo in 23 casi è stata utilizzata la tecnica di sequenziamento parallelo NGS. In generale ed è stato possibile disporre di una corrispondenza fenotipo-genotipica per poco meno di 70 malattie rare e circa 60 geni o regioni correlate. I pazienti senza diagnosi rappresentano il 10% circa del totale dei pazienti arruolati, ma di questi il 63% non dispone ancora d'indagine genetica, che è stata invece utile alla conferma diagnostica nel 53% dei pazienti con malattia nota.

L'ampliamento della popolazione di pazienti arruolati e l'applicazione mirata di tecnologie NGS potrebbe consentire nuova conoscenza più ampia conoscenza sul profilo fenotipico e genotipico delle malattie considerate.

La fascia d'età più rappresentata nel campione è quella pediatrica, sottolineando l'importanza di un valido strumenti che contribuisca all'implementazione del processo d'inquadramento diagnostico dei malati rari, in prospettiva di una miglior accessibilità alle cure ed una maggior qualità di vita.

Il lavoro di revisione della letteratura internazionale, svoltosi fin qui solo in relazione a un limitato numero di malattie rare, ha consentito la raccolta di misure di frequenza teoriche di segni clinici e reperti diagnostici, che si è mostrata utile alla creazione ancora in fase preliminare, di un'infrastruttura operativa per elaborazione di casi pazienti simulati, con profili di probabilità diagnostiche prederminate.

L'ampliamento della raccolta casi e l'integrazione dei dati di pazienti con diagnosi nota potrà consentire, all'interno di un sistema esperto, la validazione delle diagnosi simulate ed il contemporaneo ampliamento della popolazione di pazienti senza diagnosi potrà consentire, con utilizzo di reti neurali o reti di kohonen e metodologie Fuzzy la formulazione di nuove diagnosi di patologie note o l'identificazione di nuove entità nosologiche.

In conclusione il presente progetto di dottorato ha proposto una metodologia di raccolta dell'informazione clinica completa, nella sua corrispondenza fenotipo-genotipica, nell'ambito di due più ampi progetti inerenti specifiche malattie rare o gruppi di malattie rare. Tale metodologia si è dimostrata in grado di rispondere ad esigenze di applicabilità in ambito clinico, di inter-operabilità in diversi ambiti assistenziali e di ricerca clinica e di assoluta generalizzazione, ovvero di possibile utilizzo per qualsiasi patologia di cui un paziente potrebbe essere considerato affetto.

Nello specifico ambito dei pazienti senza diagnosi lo strumento si è dimostrato idoneo ai requisiti ed agli obiettivi inclusi nel progetto UnRareNet ancora in corso, con notevoli potenzialità applicative future che mirano all'ampliamento delle conoscenze sulla variabilità genomica delle malattie rare ed alla riduzione del numero dei malati senza diagnosi.

6 Bibliografia

1. Office P. Council Recommendation of 8 June 2009 on an action in the field of rare diseases. 2009; :1–4.
2. No D. 1295/1999/EC of the European Parliament and of the Council of 29 April 1999 adopting a programme of Community action on rare diseases within the Decision repealed by Decision; 2002.
3. Taruscio YKbfd, Fantini B, Taruscio D. Classification and codification of rare diseases. *Journal of Clinical Epidemiology* 2012; 65:1026–7.
4. Schieppati A, Henter JI, Daina E, Aperia A. Why rare diseases are an important medical and social issue. *The Lancet*; 2008.
5. Richesson RL, Lee HS, Cuthbertson D, Lloyd J, Young K, Krischer JP. An automated communication system in a contact registry for persons with rare diseases: Scalable tools for identifying and recruiting clinical research participants. *Contemporary Clinical Trials* 2009; 30:55–62.
6. Hayashi S, Umeda T. 35 years of Japanese policy on rare diseases. *The Lancet* 2008;
7. Lin J-D, Lin L-P, Hung W-J. Reported numbers of patients with rare diseases based on ten-year longitudinal national disability registries in Taiwan. *Research in Developmental Disabilities* 2013; 34:133–8.
8. Anderson M, Elliott EJ, Zurynski YA. Australian families living with rare disease: experiences of diagnosis, health services use and needs for psychosocial support. *Orphanet Journal of Rare Diseases* 2014 9:1 2013; 8:1–1.
9. la Paz de MP, Villaverde-Hueso A, Alonso V, János S, Zurriaga Ó, Pollán M, Abaitua-Borda I. Rare Diseases Epidemiology Research. In: *Rare Diseases Epidemiology*. Dordrecht: Springer Netherlands; 2010. pages 17–39.
10. Rath A, Olry A, Dhombres F, Brandt MM, Urbero B, Aymé S. Representation of rare diseases in health information systems: The orphanet approach to serve a wide range of end users. *Hum Mutat* 2012; 33:803–8.
11. Groft SC. A past with uncertainty, a future with hope – rare disease day 2014 from a USA perspective. *Orphanet Journal of Rare Diseases* 2014 9:1 2014; 9:1.

12. Ayme S, Rodwell C. The European Union Committee of Experts on Rare Diseases: three productive years at the service of the rare disease community. *Orphanet J Rare Dis* 2014;
13. Baldovino S, Menegatti E, Modena V. Piedmont and Aosta Valley inter-regional network in the context of the Italian National Network for rare diseases. *Blood transfusion* 2014;
14. Mazzucato M, Pozza LVD. A population-based registry as a source of health indicators for rare diseases: the ten-year experience of the Veneto Region's rare diseases registry. *Orphanet J Rare ...* 2014;
15. Donnart A, Viollet V, M R-T. [Rare diseases, definitions and epidemiology]. *Soins Pédiatrie* 2012; :14–6.
16. Guillem P, Cans C, Robert-Gnansia E. Rare diseases in disabled children: an epidemiological survey. *Archives of disease in ...* 2008;
17. Greenlee RT, Goodman MT, Lynch CF. The occurrence of rare cancers in US adults, 1995–2004. ... *health reports* 2010;
18. SHEN T, LEE A, SHEN C, LIN CJ. The long tail and rare disease research: the impact of next-generation sequencing for rare mendelian disorders. *Genet Res* 2015; 97:e15.
19. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nature Publishing Group* 2013; 14:681–91.
20. Riggs ER, Wain KE, Riethmaier D, Savage M, Smith-Packard B, Kaminsky EB, Rehm HL, Martin CL, Ledbetter DH, Faucett WA. Towards a Universal Clinical Genomics Database: The 2012 International Standards for Cytogenomic Arrays Consortium Meeting. *Hum Mutat* 2013; 34:915–9.
21. Robinson PN, Mundlos S. The Human Phenotype Ontology. *Clinical Genetics* 2010; 77:525–34.
22. Aymé S, Rath A, Bellet B. WHO International Classification of Diseases (ICD) Revision Process: incorporating rare diseases into the classification scheme: state of art. *Orphanet Journal of Rare Diseases* 2014 9:1 2010; 5:P1.
23. P L, C M, A R, Le Mignot L, E D, Ben Said M, P JJ, L T, G B, E B-L, et al. CEMARA an information system for rare diseases. *Studies in Healt Technology and Informatics* 2010; :481–5.

24. rob. Microsoft Word - ICD Revision Manual for TAG and Workgroups.docx. 2008; :1–17.
25. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. REPO R TThe Human Phenotype Ontology: A Tool for Annotating and Analyzing Human Hereditary Disease. *The American Journal of Human Genetics* 2008; 83:610–5.
26. Köhler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Bailleul-Forestier I, Black GCM, Brown DL, Brudno M, Campbell J, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Research* 2014; 42:D966–74.
27. Cornet R, de Keizer N. Forty years of SNOMED: a literature review. *BMC Med Inform Decis Mak* 2008; 8:S2.
28. Lee D, de Keizer N, Lau F, Cornet R. Literature review of SNOMED CT use. *J Am Med Inform Assoc* 2014; 21:e11–9.
29. Andrews JE, Richesson RL, Krischer J. Variation of SNOMED CT Coding of Clinical Research Concepts among Coding Experts. *J Am Med Inform Assoc* 2007; 14:497–506.
30. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research* 2014; 42:D980–5.
31. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Research* 2005; 33:D514–7.
32. Hamosh A, Scott AF, Amberger J, Bocchini C, Valle D, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Research* 2002; 30:52–5.
33. Lindblom A, Robinson PN. Bioinformatics for Human Genetics: Promises and Challenges. *Hum Mutat* 2011; 32:495–500.
34. Amberger J, Bocchini C, Hamosh A. A new face and new challenges for Online Mendelian Inheritance in Man (OMIM®). *Hum Mutat* 2011; 32:564–7.
35. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Research* 2016; 44:D862–8.

36. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, Altman RB, Klein TE. Pharmacogenomics Knowledge for Personalized Medicine. *Clin Pharmacol Ther* 2012; 92:414–7.
37. Hamosh A, Sobreira N, Fong JH. PhenoDB: A New Web-Based Tool for the Collection, Storage, and Analysis of Phenotypic Features. *Hum Mutat* 2013; :566–71.
38. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* 2014; 42:D7–17.
39. F S. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci* 1977; 74:5463–7.
40. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. *Trends in Genetics* 2014; 30:418–26.
41. Xuan J, Yu Y, Qing T, Guo L, Shi L. Next-generation sequencing in the clinic: Promises and challenges. *Cancer Letters* 2013; 340:284–95.
42. Danielsson K, Mun LJ, Lordemann A, Mao J, Lin C-HJ. Next-generation sequencing applied to rare diseases genomics. *Expert Review of Molecular Diagnostics* 2014; 14:469–87.
43. Rabbani B, Mahdieh N, Hosomichi K, Nakaoka H, Inoue I. Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. *J Hum Genet* 2012; 57:621–32.
44. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2009; 42:30–5.
45. George RA, Smith TD, Callaghan S, Hardman L, Pierides C, Horaitis O, Wouters MA, Cotton RGH. General mutation databases: analysis and review. *Journal of Medical Genetics* 2008; 45:65–70.
46. Plazzer J-P, Macrae F. DNA Variant Databases: Current State and Future Directions. In: Trent R, editor. *Clinical Bioinformatics*. New York, NY: Springer New York; 2014. pages 263–73.
47. Povey S, Aqeel AI, Cambon-Thomsen A, Dagleish R, Dunnen den JT, Firth HV, Greenblatt MS, Barash CI, Parker M, Patrinos GP, et al. Practical guidelines addressing ethical issues pertaining to the curation of human locus-specific variation databases (LSDBs). *Hum Mutat* 2010; 31:1179–84.

48. Howard HJ, Beaudet A, Gil-da-Silva Lopes V, Lyne M, Suthers G, Van den Akker P, Wertheim-Tysarowska K, Willems P, Macrae F. Disease-specific databases: Why we need them and some recommendations from the Human Variome Project Meeting, May 28, 2011. *Am J Med Genet* 2012; 158A:2763–6.
49. Peterson TA, Doughty E, Kann MG. Towards Precision Medicine: Advances in Computational Approaches for the Analysis of Human Variants. *Journal of Molecular Biology* 2013; 425:4047–63.
50. Gray KA, Yates B, Seal RL, Wright MW, Bruford EA. Genenames.org: the HGNC resources in 2015. *Nucleic Acids Research* 2015; 43:D1079–85.
51. Famiglietti ML, Estreicher A, Gos A, Bolleman J, Géhant S, Breuza L, Bridge A, Poux S, Redaschi N, Bougueleret L, et al. Genetic Variations and Diseases in UniProtKB/Swiss-Prot: The Ins and Outs of Expert Manual Curation. *Hum Mutat* 2014; 35:927–35.
52. The UniProt Consortium, Hinz U. From protein sequences to 3D-structures and beyond: the example of the UniProt Knowledgebase. *Cell Mol Life Sci* 2009; 67:1049–64.
53. Cooper DN, Ball EV, Krawczak M. The human gene mutation database. *Nucleic Acids Research* 1998; 26:285–7.
54. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NST, Abeyasinghe S, Krawczak M, Cooper DN. Human Gene Mutation Database (HGMD®): 2003 update. *Hum Mutat* 2003; 21:577–81.
55. Nair PS, Vihinen M. VariBench: a benchmark database for variations. *Hum Mutat* 2013; :42–9.
56. Bruford EA, Lush MJ, Wright MW, Sneddon TP, Povey S, Birney E. The HGNC Database in 2008: a resource for the human genome. *Nucleic Acids Research* 2008; 36:D445–8.
57. Povey S, Lovering R, Bruford E, Wright M, Lush M, Wain H. The HUGO Gene Nomenclature Committee (HGNC). *Hum Genet* 2001; 109:678–80.
58. Wain HM, Lush M, Ducluzeau F, Povey S. Genew: the Human Gene Nomenclature Database. *Nucleic Acids Research* 2002; 30:169–71.
59. Wain HM, Lush MJ, Ducluzeau F, Khodiyar VK, Povey S. Genew: the Human Gene Nomenclature Database, 2004 updates. *Nucleic Acids Research* 2004; 32:D255–7.

60. Ring HZ, Kwok P-Y, Cotton RG. Human Variome Project: an international collaboration to catalogue human genetic variation. *Pharmacogenomics* 2006; 7:969–72.
61. Cotton RGH, Auerbach AD, Axton M, Barash CI, Berkovic SF, Brookes AJ, Burn J, Cutting G, Dunnen den JT, Flicek P, et al. GENETICS: The Human Variome Project. *Science* 2008; 322:861–2.
62. Oetting WS, Robinson PN, Greenblatt MS, Cotton RG, Beck T, Carey JC, Doelken SC, Girdea M, Groza T, Hamilton CM, et al. Getting Ready for the Human Phenome Project: The 2012 Forum of the Human Variome Project. *Hum Mutat* 2013; 34:661–6.
63. HEALTH INDICATORS FOR RARE DISEASESII – Conceptual framework for the use of health indicators for monitoring quality of care. 2011; :1–30.
64. Neurology TL. The countdown for rare disease strategies is underwayEditorial. *The Lancet Neurology* 2012; 12:1.
65. Mehta A. The how (and why) of disease registers. *Early Human Development* 2010; 86:723–8.
66. Forrest CB, Bartek RJ, Rubinstein Y, Groft SC. The case for a global rare-diseases registry. *The Lancet* 2011; 377:1057–9.
67. Rare Disease Registries in Europe. 2014; :1–35.
68. Taruscio D, Gainotti S, Mollo E, Vittozzi L, Bianchi F, Ensini M, Posada M. The current situation and needs of rare disease registries in Europe. *Public Health Genomics* 2013; 16:288–98.
69. Rubinstein YR, Groft SC. Contemporary Clinical Trials. *Contemporary Clinical Trials* 2010; 31:393.
70. Rubinstein YR, Groft SC, Bartek R, Brown K, Christensen RA, Collier E, Farber A, Farmer J, Ferguson JH, Forrest CB, et al. Creating a global rare disease patient registry linked to a rare diseases biorepository database: Rare Disease-HUB (RD-HUB). *Contemporary Clinical Trials* 2010; 31:394–404.
71. Licht C, Ardissino G, Ariceta G, Cohen D, Cole JA, Gasteyger C, Greenbaum LA, Johnson S, Ogawa M, Schaefer F, et al. The global aHUS registry: methodology and initial patient characteristics. *BMC Nephrology* 2015; :1–8.
72. Maddern GJ, Smith JA, Babidge W. Development of clinical-quality registries in Australia: the way forward. ... *journal of Australia*; 2011.

73. Wang J-B, Guo JJ, Yang L, Zhang Y-D, Sun Z-Q, Zhang Y-J. Rare diseases and legislation in China. *The Lancet* 2010; 375:708–9.
74. Zhang Y-J, Wang Y-O, Li L, Guo JJ, Wang J-B. China's first rare-disease registry is under development. *The Lancet* 2011; 378:769–70.
75. Taruscio D, Kodra Y, Ferrari G. The italian national rare diseases registry. *Blood transfusion* 2014; :s606–13.
76. Taruscio D, Mollo E, Gainotti S, Posada de la Paz M, Bianchi F, Vittozzi L. The EPIRARE proposal of a set of indicators and common data elements for the European platform for rare disease registration. *Arch Public Health* 2014; 72:35.
77. charlotte. PART I: OVERVIEW OF RARE DISEASE ACTIVITIES IN EUROPE. 2012; :1–77.
78. Vittozzi L, Gainotti S, Mollo E, Donati C, Taruscio D. A model for the European platform for rare disease registries. *Public Health Genomics* 2013; 16:299–304.
79. Ministero della Salute. 2011; :1–28.
80. tg01. Fascicolo Sanitario Elettronico:come funziona, aspetti organizzativo-gestionali e infrastrutturali.Attuazione del FSE nella Regione Emilia-Romagna. 2013; :1–33.
81. Archer N, Fevrier-Thomas U, Lokker C, McKibbin KA, Straus SE. Personal health records: a scoping review. *J Am Med Inform Assoc* 2011; 18:515–22.
82. Ahmadi M, Jeddi FR, Gohari MR, Sadoughi F. A Review of the Personal Health Records in Selected Countries and Iran. *J Med Syst* 2010; 36:371–82.
83. M W, K D. Dreams and Nightmares: Practical and Ethical Issues for Patients and Physicians Using Personal Health Records. *Journal of law, medicine and ethics* 2010; :64–73.
84. Kole A, Faurisson F. Rare Diseases Social Epidemiology: Analysis of Inequalities. In: la Paz de MP, Groft SC, editors. *Rare Diseases Epidemiology*. Dordrecht: Springer Netherlands; 2010. pages 223–50.
85. Tiftt CJ, Adams D. The National Institutes of Health undiagnosed diseases program. *Current opinion in pediatrics* 2014; 26:623–33.

86. Gahl WA, Markello TC, Toro C, Fajardo KF, Sincan M, Gill F, Carlson-Donohoe H, Gropman A, Pierson TM, Golas G, et al. The National Institutes of Health Undiagnosed Diseases Program: insights into rare diseases. *Genet Med* 2011; 14:51–9.
87. Gahl WA, Tiffit CJ. The NIH Undiagnosed Diseases Program Lessons Learned. *journal of american medicine association* 2011; 305:1904.

Indice delle figure

Figura 1: Schermata principale del modulo anagrafica parte a	93
Figura 2: Schermata principale del modulo "Albero delle malattie"	94
Figura 3: Schermata principale del modulo "Piano Terapeutico Assistenziale"	94
Figura 4: Schermata principale del modulo "Controlli"; possibilità di filtrare le caratteristiche fenotipiche.....	95
Figura 5: Modulo Controlli; apertura schermata secondaria di compilazione di altre caratteristiche fenotipiche e dei trattamenti pregressi rilevanti	95
Figura 6: Modulo Controlli; sotto sezioni con entità principali, descrittive il fenotipo	95
Figura 7: Modulo Controlli; sotto sezione con entità principali segni	96
Figura 8: Modulo Controlli; schermata secondaria di ricerca segni nel thesaurus a logica gerarchica.....	96
Figura 9: Modulo Controlli; sotto sezione comorbidità	97
Figura 10: Modulo Controlli; sotto sezione genotipo.....	97
Figura 11: Distribuzione media per paziente delle entità principali.....	101
Figura 12: Andamento temporale dell'arruolamento pazienti nel corso del progetto ..	102
Figura 13: Distribuzione dei casi arruolati da ciascun CA	102
Figura 14: Numerosità dei pazienti per gruppi di malattie rare correlate.....	104
Figura 15: Distribuzione dei pazienti per fasce di età	105
Figura 16: Numerosità dei pazienti senza e con una o più indagini genetiche.....	105
Figura 17: Distribuzione numerica pazienti per assenza o presenza di diagnosi nota .	107
Figura 18: Distribuzione percentuale dei pazienti senza diagnosi per presenza o assenza di indagine genetica	108
Figura 19: Distribuzione percentuale dei pazienti con diagnosi nota per presenza o assenza di indagine genetica.....	108

Indice delle tabelle

Tabella 1: riepilogo numero totale di entità principali e secondarie e di relazioni contenute nel database relazionale	100
Tabella 2: Numero totale di entità principali riportate per ciascuno dei pazienti del campione.....	100
Tabella 3: Pazienti arruolati dai singoli CA nel corso del progetto.....	102
Tabella 4: Entità nosologiche maggiormente rappresentate per numero di pazienti diagnosticati, tra quelle contenute nel database relazionale	103
Tabella 5: Numerosità pazienti per singole tipologie di indagine genetica specifica ..	106
Tabella 6: Tecniche diagnostiche di indagine genetica molecolare	107
Tabella 7: Numero di casi per patologia rara diagnosticati con indagini di genetica molecolare	107
Tabella 8: Frequenze percentuali teoriche per alcuni segni clinici principali della acidemia metilmalonica tratti da Horster et al. 2012.....	109
Tabella 9: Percentuali di frequenza teoriche per alcuni segni clinici principali della Sindrome di Leigh tratte da Safou et al. 2014	110