Fundación InstitutoRoche 20 aniversario

Anticipating reports

PRECISION PHENOTYPING

Observatorio de Tendencias | EN LA MEDICINA DEL FUTURO

Anticipating report coordinated by:

Pablo Lapunzina
Full Professor of Human Genetics and Head of Research group at Institute of Medical and Molecular Genetic (INGEMM) of La Paz University Hospital (IdiPaz) and Scientific Director at CIBERER.

Collaborating experts:

Enrique Galán
Head of Pediatric Department at the Maternal and Child Hospital Regional University Hospital Infanta Cristina, in Badajoz. Head of Clinical Linked Groupo 23/ER/2 at CIBERER. Professor of Pediatrics and director of the Department of Biomedical Sciencies at the University of Extremadura.

Jair Antonio Tenorio
Principal investigator at the Institute of Medical and Molecular Genetic at La Paz University Hospital.

Advisory committee at the Observatory of Trends in the Medicine of the Future:

Joaquín Arenas
Director of the Research Institute at 12 de Octubre University Hospital (i+12).

Ángel Carracedo
Director of the Galician Public Foundation of Genomic Medicine (Galician Health Service) and Coordinator of the Genomic Medicine Group at the University of Santiago de Compostela (CIBERER).

Pablo Lapunzina
Full Professor of Human Genetics and Head of Research group at Institute of Medical and Molecular de Genetic (INGEMM) at La Paz University Hospital (IdiPaz) and Scientific Director at CIBERER.

Fernando Martín-Sánchez
Deputy Manager of the Medical Informatics, Digital Stratety and Innovation Area at La Paz University Hospital.

www.institutoroche.es

How to reference this report: *Fundación Instituto Roche. Anticipating Report Precision Phenotyping. 2024.*

With the methodological support of Ascendo Sanidad&Farma

Content

PRESENTATION

The Anticipating Reports, elaborated within the framework of the Observatory of Trends in the Medicine of the Future fostered by Roche Institute Foundation, are intended to contribute to the generation and sharing of advances in areas of emerging knowledge related to Personalized Precision Medicine and that will be part of the Medicine of the Future.

The Observatory has an Advisory Committee of experts formed by Dr. Ángel Carracedo, Dr. Joaquín Arenas, Dr. Pablo Lapunzina, and Dr. Fernando Martín-Sánchez. Their functions include the selection of topics addressed in these reports, the identification of experts, and the validation of content.

This report on **"Precision Phenotyping"** is coordinated by **Dr. Pablo Lapunzina**, and has been elaborated with the participation of experts **Dr. Enrique Galán** y **Dr. Jair Antonio Tenorio**.

Dr. Pablo Lapunzina holds a degree and doctorate in Medicine and Surgery from the University of Buenos Aires. He completed his residency training in pediatrics at the Children's Hospital Dr. Ricardo Gutiérrez in Buenos Aires, later serving as Chief Resident and Instructor. Currently, he is a Full Professor of Human Genetics and Head of the Research Group at the Institute of Medical and Molecular Genetics of the La Paz University Hospital, Madrid, and Scientific Director of CIBERER. He holds a Master's degree in Molecular Genetics and a Master's in Hospital Management. Dr. Lapunzina specializes in embryo-fetal medicine and is an expert in Artificial Intelligence from MIT (Massachusetts Institute of Technology in Boston, USA, 2021). He is a member of CIBERER and the European reference network ITHACA (Intellectual disability, TeleHealth, Autism and Congenital Anomalies). He is the author of over 300 articles, 35 chapters, and 14 books. His work has focused on genomic disorders, overgrowth syndromes, syndromes with growth failure, and syndromes with imprinting alterations. Along with his research group and in collaboration with several international groups, he has described over 20 new diseases and syndromes, along with the identification of several disease-associated genes

Dr. Enrique Galán holds a degree and doctorate in Medicine and Surgery from the University of Extremadura. He completed his specialized training in Pediatrics at the Maternal and Child's Infanta Cristina Hospital in Badajoz and at the Hospital Clínic i Provincial in Barcelona. Subsequently, he completed his Fellowship in Clinical Genetics, Dysmorphology, and Cytogenetics at the University of South Florida. Throughout his

professional career, he has combined clinical practice with research activities, focusing on the field of Clinical Genetics and Rare Diseases. Currently, he works as the Head of the Pediatrics Department at the Materno-Infantil Infanta Cristina Hospital in Badajoz and as the head of the Clinical Group Linked 23/ER/2 of CIBERER. He is a Professor of Pediatrics at the University of Extremadura and a full member of the Academy of Medicine of Extremadura, in addition to holding other academic positions throughout his career. He is also an active member of scientific societies and scientific and professional representation, having served as president of the Spanish Society of Clinical Genetics, as well as a member and vocal of the Spanish Association of Pediatrics, vocal of the Pediatrics specialty at the Ministry of Health, vice president of the Society of Pediatrics of Western Andalusia and Extremadura, and president of the Extremadura Society of Rare Diseases.

Dr. Jair Antonio Tenorio holds a degree in Biology from the Autonomous University of Madrid, where he completed a Master's in Pharmacological Research and later obtained his doctorate in Molecular Biosciences. Currently, he works as a Principal Investigator at the Institute of Medical and Molecular Genetics of La Paz University Hospital and is an Associate Professor in the Faculty of Medicine at the CEU San Pablo University. His professional career has focused on the study of Rare Genetic Diseases, including the application of molecular techniques, massive sequencing, and bioinformatic tools for variant analysis, prioritization, report preparation, and the application of machine learning techniques in genomics. His research areas include the analysis of next-generation DNA sequencing data, whole exome and genome sequencing, the characterization of new genetic disorders, and the study of Pulmonary Hypertension, where he specialized during a visiting professorship at Stanford University. He combines his research activities with efforts to implement genomic medicine in clinical practice through the diagnosis of genodermatoses, neurogenetics, and the study of pharmacogenetic drugs for pancreatic cancer. Additionally, he is a member of the expert committee of the Innovative Public Procurement project of the Community of Madrid, a member of the Pulmonary Hypertension expert committee of the ClinGen Consortium, a member of the Pulmonary Hypertension working group of the International Consortium of Genetic Studies in Pulmonary Hypertension, and is part of CIBERER and the European network ITHACA.

EXECUTIVE SUMMARY

The study and approach to diseases generally start with the observation of observable characteristics, signs, and symptoms presented by patients, which is known as the phenotype. However, with the development and advancement of Personalized Precision Medicine, and the availability of large amounts of data from different sources, this approach becomes obsolete.

In this context, precision phenotyping emerges, aimed at deeply characterizing the states of health and disease with the goal of understanding the relationships that exist between the various factors that affect their development and the features, signs, and symptoms that occur in an individual. Precision phenotyping aims to create a unique, coded, and internationally standardized terminology to define and represent, in as much detail as possible, the characteristics of the states of health and disease in a structured, accessible, and understandable manner, so that it can be managed and interpreted at both human and computational levels.

Furthermore, historically, phenotyping has been associated with bias related to the knowledge and training of clinicians, as well as the healthcare context of each country. Therefore, precision phenotyping seeks to objectify the analysis of the phenotype to make it accessible and universal, through the use of terminologies and ontologies, such as the Human Phenotype Ontology.

All of this is expected to contribute, in the first instance, to making more precise diagnoses, understanding the determinants of diseases, improving therapeutic strategies, understanding the changes that occur in the phenotype of diseases over time and the underlying mechanisms, all while optimizing the overall healthcare process.

However, the development of precision phenotyping is still in its early stages, to generate the knowledge and information necessary for adequate implementation and for standardizing and universalizing the terminologies and ontologies used for this purpose. Therefore, in view of the future implementation of precision phenotyping, it will be necessary to address a series of challenges of different kinds, such as the lack of standards for conducting precision phenotyping, accessibility to relevant medical information for phenotyping, obsolescence of technological infrastructures that, in turn, hinders integration between different levels of care, or the technological training of professionals.

INTRODUCTION

In the last few decades, the development of genomics, and above all, the conclusion of the Human Genome Project[a] in 2023, has allowed for the understanding, characterization, and comprehension of the genetic information of the human being. From this knowledge, associations have been established between genes and the functions they perform, as well as between genetic alterations and the signs, symptoms, and medical findings present in an individual (known as the phenotype[b]), allowing for a deeper study of diseases within the framework of Personalized Precision Medicine.[1,2]

The phenotype is the set of observable morphological and physiological traits of an individual resulting from the combination of their genome and exposome, and its study is essential for understanding the spectrum of ways in which a pathology can present.[3] Additionally, the **growing development of other omic sciences**, and the **availability of large amounts of data**, are fundamental in providing **relevant information to accurately characterize and understand diseases, as well as states of health and disease** (for further information, see the Anticipating Report Omic Sciences and Data in the era of Personalized Precision Medicine). Therefore, **defining, integrating, and analyzing human phenotype data is considered key** to advancing the understanding of human biology and genetics.

PRECISION PHENOTYPING

Precision phenotyping consists of the **precise and exhaustive description of the observable traits, signs, and symptoms that caracterizes the states of health and/or disease of individuals in a uniformly and standardizedly manner**. This description includes more general or "macro" aspects, such as clinical and demographic characteristics or findings in different medical tests, to more specific or "micro" aspects, such as omic data, genetic alterations, cellular interactions, etc. (see Figure 1). [3, 4]

[a] *Project in which the complete human genome was mapped and sequenced for the first time.*
[b] *Set of observable morphological and physiological traits of an individual with a specific genotype.*

Figure 1. Aspects that define Precision Phenotyping.

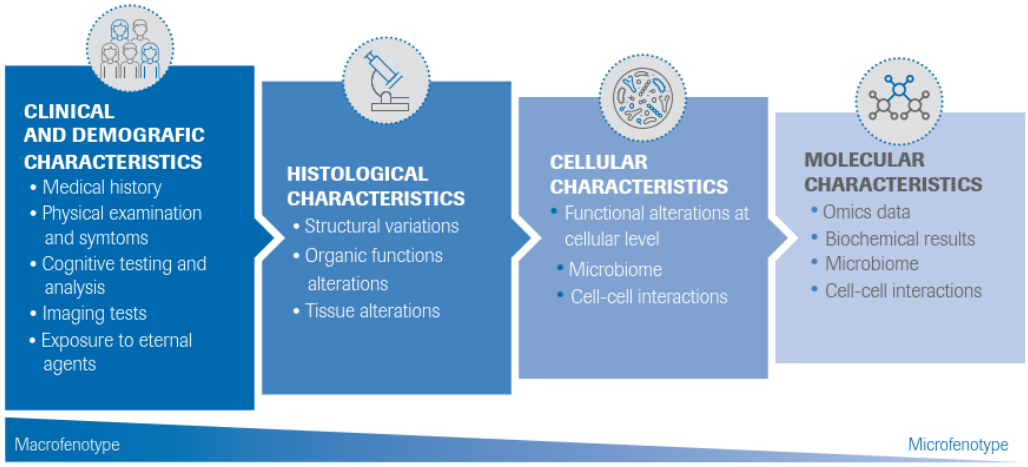

Figure 1. Aspects that define Precision Phenotyping. The definition of precision phenotyping encompasses the observation and evaluation of parameters with a high level of granularity. This process ranges from macrophenotyping, based on the analysis of observable clinical and demographic characteristics through symptomatology, medical history, and physical examination of subjects, as well as observable organic and tissue characteristics through more specific tests; to microphenotyping, which involves the analysis of cellular and molecular level characteristics, requiring laboratory analysis. Adapted from ([4]).

Precision phenotyping seeks to **employ a unique, coded, and internationally standardized terminology to define and represent, in as much detail as possible, the characteristics of states of health and disease**. Thanks to this level of precision in the description, phenotypic information could be integrated and analyzed objectively, by anyone or any computer, for application in research or clinical practice.[3,5]

In fact, in the field of biomedical research, the determination of the phenotype contributes to generating knowledge about the wide spectrum of manifestations that can be associated with a specific disease or with a state of health, or even with the effect of other factors. This knowledge can help discern whether a particular sign or symptom is associated with a specific pathology, occurs in isolation, or is an effect of, for example, pharmacological treatments. However, phenotypic descriptions made in the context of research are also imprecise and often depend on the way they are expressed by the involved research personnel.[3]

On the other hand, in clinical practice, the clinical diagnosis of diseases has generally been based on determining the phenotype of patients through medical history, physical examination, and the performance of imaging, laboratory, or psychological tests, information that is recorded in the medical record.[3] The precision in establishing this phenotype has a significant impact on the diagnosis, which influences the decision in establishing the therapeutic strategy and, therefore, on the patients' health. [2,3] However, at present, the recording of this information in the medical record is heterogeneous and

poorly standardized, as it is generally done through open text fields that must be completed by healthcare professionals.

Thus, **precision phenotyping** is proposed as a **useful tool in Personalized Precision Medicine**, as it would allow for a **greater understanding of biological variability, as well as the identification of correlations between the available patient data and the observed phenotypes of various diseases**.

Phenotype classification: ontologies y taxonomies

Precision phenotyping requires having a system that allows standardizing terminologies so that, under a single term referring to a specific characteristic (e.g. macrocephaly), all synonyms or ways of expressing such characteristic collected in medical terminology dictionaries and vocabularies (e.g. large head, increased head circumference, or megalencephaly, at times) are encompassed. This allows relevant medical information in the study of diseases to be used universally. [6,7]

For this purpose, numerous initiatives for coding have emerged, translating terms referring to a specific characteristic into a single alphanumeric code.[8] These systems are usually organized under a hierarchical structure that facilitates the analysis of information by professionals as well as computationally, allowing for the establishment of relationships between terms that facilitate their management.

Traditionally, these systems have been organized into **taxonomies** (the science of classification), which consist of structures that **classify the different elements within a hierarchy**, meaning that each element is classified within a broader category, allowing for the organization of large amounts of information in a systematic and comprehensible manner.[9] In a more general context, a taxonomy is a classification scheme that organizes concepts or elements into hierarchical and structured categories based on their common characteristics. In biology, it refers to the classification of organisms in an ordered system that indicates natural relationships. For example, the International Classification of Diseases (ICD) by the World Health Organization (WHO) classifies and codes diseases and health problems into major chapters such as "Neoplasms," "Infectious and Parasitic Diseases," or "Diseases of the Nervous System," which are further subdivided into sections and subsections that classify diseases based on criteria such as etiology, anatomical location, or type of disease. For instance, in the case of neoplasms, they are subdivided based on malignancy into "Malignant neoplasms," "Benign neoplasms," "Neoplasms of uncertain behavior," or "Neoplasms of unspecified nature," and within them into even more specific groups, such as by location "Malignant neoplasms of lip, oral cavity, and pharynx".[8, 10]

However, the large amounts of information generated within the framework of Personalized Precision Medicine, along with the need to advance towards precision phenotyping, have led to the evolution of these classifications into more complex systems that allow for establishing relationships between terms with different origins. In this context, **ontologies** are born, a branch of philosophy that studies the nature of being, existence, and reality. In the context of computer science and information science, an ontology is a formal representation of a set of concepts within a domain and the relationships between those concepts. Ontologies **present and define the terms to be used within a complex structure that allows for establishing relationships between elements from different groups or origins, facilitating automatic analysis** (see Figure 2).[8, 11, 12] It is used to model domain knowledge in a structured manner and facilitates interoperability between systems.

Figure 2. Illustrative example of how the Human Phenotype Ontology (HPO) is organized.

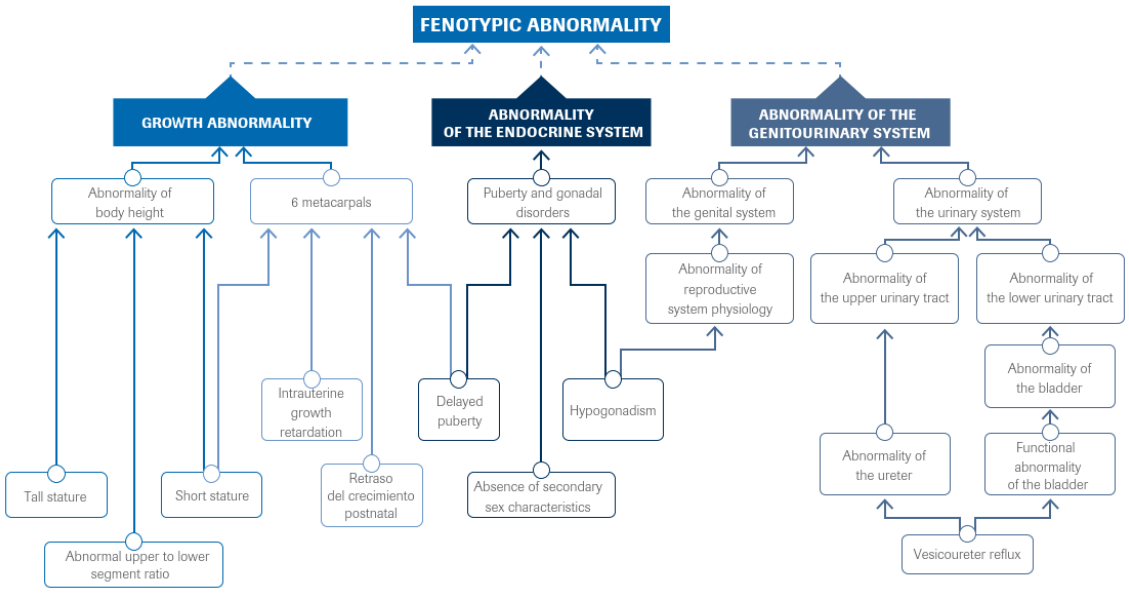

Figure 2. Illustrative example of the organization in HPO for the phenotypic abnormality with code HP:0000118. Phenotypic terms (nodes) are represented, connected by edges to generate subclasses, so that a term represents a more specific trait, sign, or symptom than its predecessor term(s). For example, "absence of secondary sexual characteristics" is a subclass of "gonadal and pubertal development anomalies," which in turn is a subclass of "endocrine system anomaly." Adapted from ([12]).

Taxonomies and ontologies are similar in that they aim to organize knowledge in a structured manner, are used to enhance the understanding and management of complex information, and are essential tools in the management and organization of knowledge in various fields. However, they differ in that ontology is more complex and detailed, while taxonomy is simpler and more linear. Ontology includes multiple types of relationships and constraints, while taxonomy primarily focuses on hierarchical relationships.

Additionally, ontology is used in the Semantic Web[c], AI, and knowledge management, while taxonomy is used in databases, information organization systems, and digital libraries.

In this way, ontologies allow, on one hand, to reach a consensus on the terms used in the medical and scientific community, and to systematize the analysis of information through the relationships between the terms that make up the ontology.[13] This, in turn, allows for making inferences and comparisons with other available sources that may have clinical relevance or be useful for research.

Below is a list of numerous initiatives focused on standardizing the terminology used to represent medical information that have emerged over the years, both for signs, symptoms, or findings (see Table 1), and for diseases (see Table 2).[8]

Table 1. Some classifications of signs, symptoms and findings commonly used in Medical Sciences, by alphabetical order

Name	Agency	Number of signs, symptoms and findigns	Type of classification
HPO (Human Phenotype Ontology)	HPO consortium	Approximately 18.000	Ontological
MedDRA-symptoms (Medical Dictionary or Regulatory Activities)	ICH (Internacional Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use)	Approximately 80.000	Hierarchical
UMLS_CUI (Unified Medical Language System - Concept Unique Identifier)	NLM (National Library of Medicine of the United States)	More tahn 2.000.000 terms	Non ontological

[c] *Extension of the World Wide Web in which the meaning or semantics of information and services are defined, allowing for a more effective understanding and processing of information, making it readable to both humans and machines.*

Table 2. Some classifications of diseases commonly used in Medical Sciences, by alphabetical order. (Part 1/2)

Name	Agency	Number of diseases	Type of classification
DO (Disease Ontology)	Genomic Sciences Institute of Maryland University	Aprpoximately 12.000	Ontological
eRAM (Encyclopedia of Rare Disease Annotations for Precision Medicine)	Biomedical Sciences Institute of East China Normal University	15.942 rare diseases	Non ontological
GARD (Genetic and Rare Disease)	NationalCenter for the Advance of Traslational Sciences of NIH (National Institute of Health of United States)	Approximately 10.000	Semi-ontological
ICD 9 (International Classification of Diseases version 9)	World Health Organization	Approximately 14.000	Hierarchical and not very ontological
ICD10 (International Classification of Diseases version 10)	World Health Organization	Approximatlye 70.000	Hierarchical and more ontological
ICD11 (International Classification of Diseases version 11)	World Health Organization	Approximately 55.000	Ontological
IND (International Nomenclature of Diseases)	AIMedicum (Artificial Intelligence in Medicine)	Approximately 53.000 diseases	Semi-Ontological
MALACARD	Weizmann Institute of Science	22.960	Non ontoogical, with subdivisions
MEdDRA-Disease (Medical Dictionary for Regulatory Activities)	ICH (Internacional Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use)	Approximately 20.000	Ontological
MEDGEN (Medical Genetics)	NCBI (National Center for Biotechnological Information)	Approximately 13.000 diseases	Ontological

Table 2. Some classifications of diseases commonly used in Medical Sciences, by alphabetical order. Part (2/2)

Name	Agency	Number of diseases	Type of classification
MonDO ID (Monarch Disease Ontology identifiers)	NIH Monarch Initiative	52.683	Ontological
NCI (National Cancer Institute)	NCI-NIH (Cancer National Institute of NIH)	Approximately 15.000 terms	Ontological
NINDS (National Institute of Neurological Disorders and Stroke)	NINDS-NIH (National Institute of Neurological and Stroke Disorders of NIH)	Not available	Hierarchical
OMIM (Online Mendelian Inheritance in Man)	John Hopkins Hospital and University	Approximately 7.400	Non ontological, with subdivisions and phenotypic series
Orphanet	INSERM (Nacional Institute of Medicine and Health Research of France)	6.346	Semi-ontológica
POSSUM-OSSUM (Pictures of Standard Syndromes and Undiag-nosed Malformations)	Servicio de Genética Clínica Victorian, del Instituto de Investigación Infantil Murdoch	Approximately 5.000 syndromes	Non ontological
Skeletal Dyslasias NOS Nomenclature	International Consortium Skeletal Diseases, group of experts	771 diseases	Non ontological classified in disease families
SNOMED (Systematized Nomenclature of Medicine) **Concept Code**	International SNOMED	More of 300.000 terms	Hierarchical and more ontological
SNOMED-CT (Clinical Terms)	International SNOMED	More than 300.000 terms	Hierarchical and more ontological
SNOMED FSN (Fully Specified Name)	International SNOMED	More than 300.000 terms	Hierarchical and more ontological

ADVANCES IN PRECISION PHENOTYPING

The development of Personalized Precision Medicine, with the growing knowledge generated through omics sciences, together with technological and computational advances, has brought numerous breakthroughs in the study and understanding of the underlying mechanisms of health and disease.

The availability of high-throughput genomic and multi-omic analyses, advanced informatics, and the analysis of large amounts of data through big data and Artificial Intelligence, are contributing to the generation of increasing knowledge about the influence of different factors on phenotype, such as genotype, environmental determinants, or lifestyle. In this sense, precision phenotyping is positioned as an **essential element to establish these associations between health determinants**, such as genotype and phenotype, and to **advance in Personalized Precision Medicine through a more precise classification of signs and symptoms**, and therefore diseases, with the ultimate goal of improving patient health. [6,7]

However, the integration of phenotypic information within the framework of Personalized Precision Medicine presents a challenge in itself and requires the **adoption of new coding initiatives to advance towards the universal implementation of precision phenotyping**. Below, we present some of these strategies that are expected to enable progress towards the implementation and use of precision phenotyping in both medical practice and future research.

Signs, symptoms and findings codification in Personalized Precision Medicine: HPO ontology

Health systems globally generally use medical vocabularies or terminologies for the collection of relevant information for the clinical history, such as ICD or SNOMED-CT. These vocabularies or terminologies are standardized representations of the terms to be used in the clinical environment and are oriented towards the final diagnosis of the patient rather than clinical findings (symptoms and signs). However, most of these vocabularies are not designed to be integrated with other data sources such as genomic data, greatly limiting the ability to capture the heterogeneity of disease phenotypes.[6]

Additionally, there are multiple terminologies developed, and not all countries, health systems, and centers use the same ones. This, coupled with the fact that most of these terminologies are not interoperable with each other, creates multiple barriers for the use of information collected in medical records regarding disease phenotypes for research purpose. [6]

On the other hand, ontologies are systems that provide more detailed and organized representations that can be computationally processed to achieve their integration and interoperability with other data sources, such as databases with graphs[d], establishing relationships between the data. [6]

Although many ontologies have been developed in recent decades for this purpose, the **Human Phenotype Ontology (HPO)** initiative is the most widely used for the coding of medical signs, symptoms, and findings. HPO was established in 2008 with the initial goal of **integrating all phenotypic information related to the anomalies presented by monogenic diseases**.[14] Since then, the project has evolved and expanded its scope, aiming not only to **integrate all phenotypic information contained in scientific and clinical databases on monogenic diseases**, but also on **common diseases**, through the mapping of terms contained in other medical-scientific ontologies and vocabularies including ORDO, MedDRA, UMLS, or SNOMED, among others. Additionally, HPO seeks to represent this information under a single code and structure that, through the use of bioinformatic tools, allows the analysis of human diseases, their phenotypes, and the establishment of relationships with other data.[5,15]

Recently, within the framework of this initiative, the HPO Internationalization Effort (HPOIE) alliance has been launched with the aim of coordinating efforts to break down linguistic barriers and harmonize the terms used in international vocabularies and terminologies, through the creation of working groups for translation. In the case of Spain, the translation of HPO began in 2013 and is implemented in numerous institutions and Clinical Genetics departments of hospitals nationwide, positioning itself as the standard terminology for information exchange in national research projects such as the IMPaCT (after the Spanish initials, Precision Medicine Infraestructure associated with Science and Technology) project or in the Biomedical Research Center for Rare Diseases.[5,15]

Computational advances in precision phenotyping

The large amounts of data generated in the context of Personalized Precision Medicine have driven the design and improvement of computational tools and infrastructures to increase the capacity for information analysis.

The success in the development and implementation of precision phenotyping will inevitably depend on the **development of computer and computational tools that**

[d] Mathematical representation of data used in some databases

allow the capture, storage, and exchange of phenotypic data, as well as its integration with other data sources. [6,7]

Thus, precision phenotyping requires adopting a computational approach and making use of available tools to enhance the use of phenotypic information in the study of diseases and in future medicine. In this context, a series of computational advancements have been identified that are expected to have a significant impact on the development of precision phenotyping and its future application:

- **Computational phenotyping.** Computational phenotyping refers to the transformation of phenotypic descriptions documented in medical records into classified codes or terms, allowing them to be processed and interpreted by a computer to extract valuable information for clinical and investigative practice.[6] For example, through the HPO coding system that establishes unique codes to define the signs and symptoms presented by diseases.[5]
- **Phenotypes storage**. Phenotype storage. To facilitate the storage and exchange of phenotypic data, within the framework of the HPO initiative, the Global Alliance for Genomics and Health (GA4GH) has created tools such as "Phenopackets". These are representations of the relevant clinical data of an individual that link phenotypic information with information about diseases, genetic information, and information about the patient in a format that uses standard ontologies to ensure interoperability between different sources, simplify data mining, and enable automatic reasoning. Additionally, it is compatible with different programming languages, making it an easily adaptable tool.[17,18]
- **Natural Language Processing (NLP).** The collection of phenotypic information in medical records and scientific publications is often not encoded. Therefore, the use of tools for natural language processing is essential for the extraction of medical text and for establishing semantic relationships with the terms and codes recorded in ontologies.[7]
- **Machine learning for phenotype analysis.** The structuring of phenotypic information in ontologies allows the application of complex Artificial Intelligence algorithms to make inferences and associations by combining phenotypic data with genomic data and data from other sources for various objectives and applications, such as differential diagnosis, personalization of the diagnostic process, or the association of the phenotype with its genetic origin.[15]

Ultimately, it is expected that the **development of computational tools and the improvement of the capabilities of information infrastructures for the analysis and storage of phenotypic data** will contribute to **enhancing the development of precision phenotyping** and, thereby, expanding its potential applications to clinical practice and research.

PRECISION PHENOTYPING APPLICATIONS

The main objective of precision phenotyping is to **create a unique, encoded, and standardized international terminology that represents medical information in a structured, accessible, and understandable manner**, allowing it to be **managed and interpreted by any person or computer in any healthcare system or context**. Among other things, this will enable more precise classifications of diseases based on phenotype, understanding the underlying mechanisms of different disease phenotypes, and defining Personalized Precision Medicine strategies tailored to each phenotype

Below are some examples of the **potential application of precision phenotyping and the derived knowledge as a support tool in research** for the study of diseases and their underlying mechanisms, as well as potential applications **in clinical practice** that are in the research or implementation phase. It is worth noting that, while precision phenotyping has been notably developed in the study of rare and monogenic diseases, it is foreseeable that the knowledge generated through precision phenotyping will enable an understanding of heterogeneity and the application of this knowledge to the care of common diseases as a whole.

Ultimately, precision phenotyping using ontologies such as HPO has the potential to transform all aspects of medicine, from diagnosis and treatment to research and public health, promoting a more precise, personalized, and efficient approach to healthcare.

Application of precision phenotyping in research

Precision phenotyping opens up a range of opportunities to improve research in the field of Personalized Precision Medicine and to promote the translation of generated knowledge into clinical practice.

- **Study of the functional implications of genes.** Despite advances in human genetics, the functional implications of many genes on disease phenotype are unknown. Therefore, the availability of precision phenotyping and coding allows for phenotypes to be compared with those of other species for which more information is available in this regard. The aim is to map the phenotypes presented by humans with the phenotype of another species and analyze whether any alteration in a gene directly related to that phenotype has been described in that species. For example, establishing correlations between the phenotypes recorded in HPO and those recorded in the mammalian phenotype ontology (MP) through automated tools like MouseFinder. [5,19]

- **Optimization of clinical trials through the identification of new therapeutic targets.** For example, HPO can help select patients with specific phenotypes for clinical trials, increasing the likelihood of study success. This, in turn, allows for the detailed analysis of the phenotypes of patients included in these studies, revealing new biological pathways that may be targeted by innovative therapies.
- **Identification of population trends and patterns**. Monitoring population health serves to understand the epidemiology of diseases. In this sense, the collection and analysis of large-scale phenotypic data can help increase knowledge about the factors affecting population health, generating new opportunities for prevention and the early identification of at-risk phenotypes, ultimately contributing to the implementation of preventive measures before diseases fully manifest.

Application of precision phenotyping in disease diagnosis

The diagnosis is a key stage in the management of any disease. The level of precision achieved in the diagnosis largely determines its subsequent management. In general, disease diagnosis starts with phenotyping. In this sense, precision phenotyping can contribute to improving the diagnosis and, consequently, the management of diseases in general.

The main application of precision phenotyping in disease diagnosis is for **differential diagnosis**. So far, differential diagnosis, understood as the selection of different diseases as possible diagnoses based on a set of clinical characteristics presented by a patient, has largely depended on the clinician's experience, the description of phenotypic anomalies, or the availability of different tests and results necessary to make an appropriate diagnosis. In this regard, the availability of phenotypic information in ontologies facilitates a standardized use of terms and descriptions to carry out precision phenotyping, which ultimately allows for a more precise differential diagnosis.

In the context of the HPO initiative, statistical models have been developed, such as Phenomizer, which allow for ranking possible diagnoses according to statistical significance (p-value) based on metrics derived from semantic analogies. These types of tools serve as support for diagnosis by establishing a list of possible diagnoses ranked by probability, which can also help healthcare professionals determine if additional tests are necessary for each patient in order to extract more phenotypic characteristics that may rule out one diagnosis or another, in other words, for the differential diagnosis of diseases.[20] Additionally, through the comparison of detailed phenotypes, it will be

possible to differentiate between diseases with similar clinical presentations, thus improving diagnostic precision.

Another example of the application of ontologies, specifically HPO, in diagnosis, is for the detection of rare diseases that may be difficult to diagnose using conventional methods, thanks to their ability to provide detailed descriptions of a patient's signs and symptoms.

Ultimately, precision phenotyping will support clinical decision support systems by applying tools for AI-assisted diagnosis, which can provide diagnostic suggestions and therapeutic recommendations based on complex phenotypic data, allowing for personalized medical care.

Application of precision phenotyping in disease treatment and monitoring

As Personalized Precision Medicine advances, new treatments are increasingly tailored to individual characteristics. Therefore, as mentioned in the previous section, the precision of the diagnosis largely determines the approach and, therefore, the treatment of diseases.

In this way, precision phenotyping allows for correlating specific phenotypic characteristics with certain responses to treatments, which in turn can contribute to designing more effective therapeutic strategies for each patient.

It is worth mentioning at this point that the human phenotype is not static, but evolves over time based on the patient's health and disease states, as well as the interaction with treatments for different diseases. For example, in the case of chronic diseases such as cancer or cardiovascular diseases, the prescription of treatments is not only influenced by the basal phenotype of the disease, but by a complex set of interactions that occur throughout the entire process, including evolution and progression, surgical interventions, or other previous treatments.

In this sense, the study of changes in phenotype over time, known as temporal phenotyping, in real clinical settings (Real World Evidence) can be relevant when studying the evolution of diseases, making more precise real-time adjustments based on the evolution of symptoms, monitoring medical activity in addressing them, and evaluating possible improvements in clinical practice guidelines. [22,23]

Application of precision phenotyping in the training of healthcare professionals

As mentioned in earlier sections of this document, phenotyping has, until now, been based on subjective descriptions, largely depending on the training and experience of healthcare professionals.

In this regard, precision phenotyping through ontologies such as HPO, which provide a standardized and universal language, can contribute to reducing differences in phenotyping by objectifying the collected information. To achieve this, expanding the use of precision phenotyping beyond clinical and research practice, implementing it from the training and education of professionals, is a strategy that will, in the future, improve clinicians' knowledge and understanding of the phenotypic diversity of diseases from the early stages of their training.

CHALLENGES

Part of the success of applying genomics in the context of Personalized Precision Medicine lies in the ability to identify and understand the consequences of genomic alterations on the human phenotype. Additionally, genomics converges with a wave of new data generated by other omics sciences such as transcriptomics, metabolomics, or proteomics, among others, which also need to be addressed and understood. Therefore, interpreting the data and its relationship with the phenotype of health and disease states, with the aim of understanding the underlying molecular bases and mechanisms of diseases, is essential for the development of future medicine.[11]

In this context, the development of precision phenotyping that allows for a detailed description of the human phenotype in different health and disease states, structured and encoded in a way that can be integrated with other sources of information and managed both at a human and computational level, is a key element. However, precision phenotyping is a field that is still in development stages and faces a series of challenges of different origins in order to optimize the information that can be obtained from its study and its implementation in real clinical practice.

Technical challenges

Precision phenotyping and the knowledge derived from it, combined with the data generated in the context of Personalized Precision Medicine, are expected to significantly contribute to the study and understanding of diseases with the aim of improving their approach from diagnosis. However, the complexity of phenotypic information and the need for integration with omics sciences and other technological advances are associated with a series of technical challenges that affect its progress.

- **Lack of standards for conducting precision phenotyping**. There is currently no standardized guide for optimal precision phenotyping, both in terms of the number of codes to incorporate and the detail and precision of the same.
- **Integration between ontologies and medical vocabularies and terminologies**. It is necessary to correctly integrate the terms contained in medical vocabularies with the codes of the ontologies used to favor interoperability between systems. However, mapping medical vocabularies is a time-consuming task, and manual mapping is prone to errors. In this regard, initiatives are emerging to carry out this automatic mapping; however, the high semantic complexity of medical information means that, in many cases, the detail

of the descriptions and their translation into ontological codes is not captured accurately.[6]

- **Lack of integration of information systems in different healthcare systems, care levels, and centers.** In the specific case of Spain, each Autonomous Community employs different information systems, including systems for the registration of Medical Records, which may not be accessible between different care levels. This means that in most cases, the information recorded in the Medical Record is not interoperable and/or shared between communities, limiting the integration of information for clinical practice and, in particular, the establishment of a system for precision phenotyping.
- **Obsolescence of hospital and healthcare system technological infrastructures.** In general, the technological infrastructures and computers used in the healthcare field are outdated. However, the advent of new computational tools and the large amounts of data being handled require an update and increased computational capacity.

Implementation challenges

In addition to the technical challenges, there are other difficulties and barriers that hinder the implementation of precision phenotyping in clinical and research practice, as well as the implementation of other necessary technological advances to optimize the application of precision phenotyping.

- **Scarcity of recorded and coded phenotypic information**. In many cases, the diagnosis is coded, but information about the signs and symptoms, i.e., the phenotypic information relevant to the diagnosis, is recorded in free text and, in some cases, it is even voluntarily. This limits access to this type of information and makes the study of diseases challenging.[7]
- **Need for resources to translate phenotypic information into codes.** Health-care activity requires completing the medical record during consultations; however, the limited time in consultations does not allow for on-the-spot translation into codes, but rather should be done afterwards. However, the high workload and lack of dedicated resources for this translation greatly limit the system's capabilities to ultimately have coded information about the phenotype.
- **Lack of consensus on the coding system to use.** There are currently numerous coding systems used worldwide. This fact often leads to non-interoperable systems, making their management difficult. For example, in the

field of genetics, the HPO system is well-established and is the only one recognized by genetic analysis software. In the field of pharmacology and therapeutics, signs and symptoms of adverse drug effects are recorded in MedDRA terminology. Additionally, in clinical practice, patients' diseases are mostly coded and not signs and symptoms based on the ICD-10 code or with SNOMED or others.

- **Lack of accesability to relevant medical information encodes.** As mentioned throughout the document, there are numerous initiatives that encode medical information about diseases, signd, symptoms and findigns. However, not all of these systems are easily accessible. For evample, MeDRA is subjected to a suscription. This is important in order to facilitate integration between the different systems.
- **Lack of knowledge and training of professionals regarding precision phenotyping.** There is a lack of awareness about the need to provide detailed descriptions of the phenotype in order to study the origin of diseases in detail. Currently, clinical practice is based on formulating diagnostic hypotheses based on a series of symptoms and signs presented by the patient, and then establishing a treatment. However, it is necessary to have adequate phenotyping to understand the basis of the patient's diseases and to carry out more precise approaches, especially in a context where the available information about patients is increasing and becoming more specific.
- **Lack of technological training for healthcare professionals.** Scientific advances are accompanied by technological and computational advances for their implementation and optimization. However, healthcare professionals generally do not have a good background in medical informatics, which hinders the use and management of computer tools in their routine practice. Therefore, there is a need to promote a change in mindset with greater interest in informatics **in order to implement scientific advances in the most efficient way.**
- **Lack of interoperability between terminologies and codifications** of symptoms, signs, and medical findings of diseases (HPO codes) and the symptoms, signs, and medical findings of adverse drug effects (for example, in MedDRA codes). This can pose a challenge when implementing precision phenotyping in routine clinical practice.

CONCLUSIONS Y RECOMMENDATIONS

Precision phenotyping allows for a detailed and standardized description of observable traits, signs, and symptoms of health and disease states. This precision is essential both in biomedical research and in clinical practice to improve the diagnosis and treatment of diseases.

The integration of phenotypic data with genomic and other omics data is key to advancing Personalized Precision Medicine, enabling a better understanding of the molecular bases and mechanisms of diseases. In this sense, HPO stands out as a fundamental tool for the encoding and standardization of phenotypic information. This ontology facilitates interoperability and data integration across different systems and databases, which is crucial for the advancement of precision medicine.

However, there are significant challenges of different natures that will need to be addressed to promote the implementation and effective use of precision phenotyping and the HPO ontology in clinical and research practice, thus contributing to the evolution of Personalized Precision Medicine.

Recommendations

- **Promote the development and adoption of international standards** for phenotype coding, using ontologies such as HPO, to ensure interoperability and precision in the description of phenotypic data.
- **Encourage international collaboration for the harmonization of terminologies and the creation of working groups** dedicated to the translation and adaptation of these ontologies to different linguistic and cultural contexts.
- **Implement training and continuous education** programs for healthcare professionals on the importance of precision phenotyping and the use of computational and ontological tools in clinical practice.
- **Include specific modules on precision phenotyping and bioinformatics** in medical training curricula and professional development programs.
- **Modernize the technological infrastructures of healthcare** systems to support the storage, processing, and analysis of large volumes of phenotypic and genomic data.
- **Develop and implement interoperable health information systems** that allow for the integration and exchange of data at national and international levels.

- **Support research in the field of precision** phenotyping to develop new tools and methods that facilitate the coding, storage, and analysis of phenotypic data.
- **Encourage collaboration between research institutions, hospitals, and technology companies** to develop innovative solutions that address the identified technical and implementation challenges.

BIBLIOGRAFY

1. Human Genome Project. National Human Genome Research Institute. Published May 7, 2024. Accessed May 8, 2024. https://www.genome.gov/es/genetics-glossary/Proyecto-Genoma-Humano

2. Delude CM. Deep Phenotyping. The details of disease. *Nature*. 2015;527:14-15. doi:https://doi.org/10.1038/527S14a

3. Robinson PN. Deep phenotyping for precision medicine. *Hum Mutat*. 2012;33(5):777-780. doi:10.1002/humu.22080

4. Wright JT, Herzberg MC. Science for the Next Century: Deep Phenotyping. *J Dent Res*. 2021;100(8):785-789. doi:10.1177/00220345211001850

5. Köhler S, Vasilevsky NA, Engelstad M, et al. The human phenotype ontology in 2017. *Nucleic Acids Res*. 2017;45(D1):D865-D876. doi:10.1093/nar/gkw1039

6. Callahan TJ, Stefanski AL, Wyrwa JM, et al. Ontologizing health systems data at scale: making translational discovery a reality. *NPJ Digit Med*. 2023;6(1). doi:10.1038/s41746-023-00830-x

7. Weng C, Shah NH, Hripcsak G. Deep phenotyping: Embracing complexity and temporality—Towards scalability, portability, and interoperability. *J Biomed Inform*. 2020;105. doi:10.1016/j.jbi.2020.103433

8. Gkoutos G V., Schofield PN, Hoehndorf R. The anatomy of phenotype ontologies: principles, properties and applications. *Brief Bioinform*. 2018;19(5):1008-1021. doi:10.1093/bib/bbx035

9. Arp R, Smith B, Spear AD. *Building Ontologies with Basic Formal Ontology*.; 2015.

10. The International Classification fo Diseases – *10th Revision, Cllinical Modification Clínica*.; 2018. Accessed May 8, 2024. https://www.sanidad.gob.es/estadEstudios/estadisticas/normalizacion/CIE10/CIE10ES_2018_diag_pdf_20180202.pdf

11. Köhler S, Doelken SC, Mungall CJ, et al. The Human Phenotype Ontology project: Linking molecular biology and disease through phenotype data. *Nucleic Acids Res*. 2014;42(D1). doi:10.1093/nar/gkt1026

12. Xue H, Peng J, Shang X. Predicting disease-related phenotypes using an integrated phenotype similarity measurement based on HPO. *BMC Syst Biol*. 2019;13. doi:10.1186/s12918-019-0697-8

13. Roussey C, Pinet F, Kang MA, Corcho O. An introduction to ontologies and ontology engineering. In: *Advanced Information and Knowledge Processing*. Vol 1. Springer London; 2011:9-38. doi:10.1007/978-0-85729-724-2_2

14. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: A Tool for Annotating and Analyzing Human Hereditary Disease. *The American Journal of Human Genetics*. 2008;83(5):610-615. doi:10.1016/j.ajhg.2008.09.017

15. Gargano MA, Matentzoglu N, Coleman B, et al. The Human Phenotype Ontology in 2024: phenotypes around the world. *Nucleic Acids Res*. 2024;52(D1):D1333-D1346. doi:10.1093/nar/gkad1005

16. Danis D, Jacobsen JOB, Wagner AH, et al. Phenopacket-tools: Building and validating GA4GH Phenopackets. *PLoS One*. 2023;18(5 MAY). doi:10.1371/journal.pone.0285433

17. Steinhaus R, Proft S, Seelow E, Schalau T, Robinson PN, Seelow D. Deep phenotyping: symptom annotation made simple with SAMS. *Nucleic Acids Res*. 2022;50(W1):W677-W681. doi:10.1093/nar/gkac329

18. Phenopackets. Published 2016. Accessed May 14, 2024. http://phenopackets.org/

19. Chen CK, Mungall CJ, Gkoutos G V., et al. Mousefinder: Candidate disease genes from mouse phenotype data. *Hum Mutat*. 2012;33(5):858-866. doi:10.1002/humu.22051

20. Köhler S, Schulz MH, Krawitz P, et al. Clinical Diagnostics in Human Genetics with Semantic Similarity Searches in Ontologies. *Am J Hum Genet*. 2009;85(4):457-464. doi:10.1016/j.ajhg.2009.09.003

21. Phenomizer tool. Published 2009. Accessed May 14, 2024. https://compbio.charite.de/phenomizer/

22. Meng W, Ou W, Chandwani S, Chen X, Black W, Cai Z. Temporal phenotyping by mining healthcare data to derive lines of therapy for cancer. *J Biomed Inform*. 2019;100. doi:10.1016/j.jbi.2019.103335

23. Zhao J, Zhang Y, Schlueter DJ, et al. Detecting time-evolving phenotypic topics via tensor factorization on electronic health records: Cardiovascular disease case study. *J Biomed Inform*. 2019;98. doi:10.1016/j.jbi.2019.103270

2018

Anticipating report
Microbiome

Anticipating report
Personalized Preventive Medicine

Anticipating report
Systems biology

2019

Anticipating report
Bioprinting

Anticipating report
Data in the era of Personalized Precision Medicine

Anticipating report
Omic sciences

2020

Anticipating report
Advanced therapies: cell therapy and gene therapy

Anticipating report
Artificial Intelligence: legal and ethical challenges

Anticipating rerpot
Exposome

2021

Anticipating report
Farmacogenomics: the path towards the personalization of treatment

Anticipating report
Nanomedicine

Anticipating report
Epigenomics

2022

Anticipating report
Nucleome 4D

Anticipating report
Radiomics

Anticipating reports
Disease risk prediction in the era of Personalized Prediction Medicine

2023

Informe Anticipando
Artificial Intelligence applications in Personalized Precision Medicine

Anticipating report
Precision vaccines

Anticipating report
Farmacological investigation en in the era of Personalized Precision Medicine

2024

Anticipating report
Precisión phenotyping

Fundación InstitutoRoche

www.institutoroche.es