10 years ESJ
Special edition

A Focused Review of Deep Phenotyping with Examples from Neurology

Daniel B. Hier, MD
Raghu Yelugam, MS
Sima Azizi, PhD
Donald C. Wunsch III, MS
Missouri University of Science and Technology, USA

Doi:10.19044/esj.2022.v18n4p4

Submitted: 10 November 2021
Accepted: 20 January 2022
Published: 08 February 2022

Cite As:
Hier B.H., Yelugam R., Azizi S. & Wunsch D.C. (2022). *A Focused Review of Deep Phenotyping with Examples from Neurology*. European Scientific Journal, ESJ, 18 (4), 4. https://doi.org/10.19044/esj.2022.v18n4p4

Abstract

A deep phenotype is the detailed description of the observable signs and symptoms, the mode of onset, the clinical course, and the response to treatment that characterizes a human disease. With advances in high-throughput phenotyping based on natural language processing and other automated algorithms, it is possible to calculate the distances between cohorts of patients and calculate the distances between individual patients. Vector representations of phenotypes allows the quantitative characterization of disease phenotypes, helps to identify phenotypic features with the highest diagnostic value, facilitate the recognition of disease variants, and supports progress towards precision medicine. This focused review introduces the underlying concepts of deep phenotype, ontology, disease repository, disease mimic, disease chameleon, patient distance, computable phenotype, and phenomics and provides illustrative examples from neurology. In closing, future advances in the application of deep phenotyping will depend on improved methods for the high-throughput phenotyping of large numbers of patients based on the unstructured text that is held in electronic health records.

Keywords: Phenotype, ontology, precision medicine, disease, disease variant, patient similarity, patient distance, high-throughput phenotyping, electronic

health records, deep phenotyping

Introduction

The Precision Medicine Initiative utilizes clinical, genomic, metabolic, and environmental data to personalize treatment at the level of the individual patient (Collins & Varmus, 2015). The key to this effort is making the clinical phenotype of each patient computable (Tasker, 2017). The clinical phenotype is each patient's externally observable disease characteristics, including signs, symptoms, mode of onset, course progression, and response to treatment. Most of the data needed for phenotyping for precision medicine is stored in electronic health records as free text and is in a non-computable format. Unlocking the value of unstructured clinical data in electronic health records depends on the successful development of high-throughput methods to create computable phenotypes (Tasker, 2017). In this focused review, we examine the status of deep phenotyping, emphasizing examples from neurology. We define deep phenotyping as the detailed description of the signs and symptoms of a disease utilizing concepts from a suitable ontology with corresponding machine-readable codes (Robinson, 2012).

Methods

For our focused review of deep phenotyping, we searched the titles of articles in the Ovid Medline database (1946 to November 12, 2021) using the search terms detailed in Table 1. Priority was given to articles focused on the deep phenotyping of neurological diseases

Table 1: Documents retrieved by Ovid Medline database

Search Terms	**Retrieved**	**Reviewed**	**Retained**
phenomics	266	22	3
pathognomonic sign	56	2	3
phenotyping AND (deep OR high throughput)	523	35	13
disease AND (variant OR subtype)	154	17	10
patient AND (distance or similarity)	240	22	4
phenotype AND ontology	57	28	8
phenotype AND repository	76	12	6
Totals	**1372**	**138**	**53**

This review was focused on introducing important concepts such as deep phenotype, disease ontology, high throughput phenotyping, disease entities, variants, mimics, and chameleons, computable phenotypes, and patient distances. The phenotype vectors for the Parkinson patient shown in Table 2 were based on methods described in Hier et al. (2020). The disease vectors shown in Table 3 are based on data and methods described in Hier & Wunsch (2021). Inter-disease distances shown in Table 4 are cosine distances

based on the vectors in Table 3 and by methods described in Hier et al. (2020). Information gain based on feature selection for different neurologic diseases is based on data from Wunsch & Hier (2021) and the feature selection widget from Orange (Demsar et al. 2013). The MDS map in Figure 2 is based on the distances in Table 4 and the MDS widget from Orange (Demsar et al. 2013).

What is deep phenotyping?

The phenotype of a disease is the collective observable manifestations of a disease. Traditionally, the phenotype was characterized by signs (findings elicited by the physician) combined with the symptoms (patient complaints). More recently, the concept of disease phenotype has been enlarged to encompass age at onset, mode of onset (acute, subacute, or insidious), rate of progression (static, progressive, relapsing, indolent), and response to medication and therapy (good, fair, or poor) (Robinson, 2012; Delude, 2015). Some have further enlarged phenotypes to include biomarkers, radiological findings, and electrophysiological findings (Bycroft et al., 2018). We use the term phenotype (Robinson, 2012) "...as the precise and comprehensive analysis of phenotypic abnormalities in which the individual components of the phenotype are observed and described, often for scientific examination of human disease". Consider a patient with Parkinson's disease that presents with the following history (Figure 1).

> *A 75-year-old man comes to the office complaining of increasing slowness and stiffness (Figure 1). Neurologic examination shows a resting tremor in both hands, a stooped posture, cogwheel rigidity, and a festinating shuffling gait. In the past, he has demonstrated an excellent response to levodopa therapy.*

Three cardinal features of Parkinson disease are bradykinesia, resting tremor, and cogwheel rigidity (Table 1). No individual patient is a perfect match to a disease phenotype, but most with a given disease will have most of the important features of that disease (Caplan, 1982. Caplan, 2021). The signs and symptoms of this patient (Table 2) can be converted to machine-readable codes from the UMLS Metathesaurus (Bodenreider, 2004). These signs and symptoms are a close but not perfect match to the known signs and symptoms of Parkinson disease.

Figure 1: Drawing by Sir William Gowers of the patient with Parkinson disease illustrating rigidity, resting tremor, masked face, and stooped posture from his 1886 textbook *A Manual of Diseases of the Nervous System* (Work is in the public domain).

What ontologies are available for phenotyping?

Disease ontologies are carefully constructed limited vocabularies that include the terms needed to describe knowledge in a restricted domain (Gruber, 2018; Jepsen, 2009). Each entry in the ontology is called a *concept* and has a distinct meaning different from all other concepts. Most disease ontologies are organized as *subsumptive hierarchies* with more specific concepts subsumed by more general concepts. For example, in a neurological ontology (Hier & Brint, 2020), *hand weakness* is subsumed by *limb weakness*, *which* is subsumed by *motor findings*. Clinical phenotyping is the process of converting disease manifestations to concepts in a disease ontology. Disease ontologies contain disease concepts, their descriptions, and their machine-readable codes. Some disease ontologies are comprehensive and have more

concepts than needed to phenotype a typical patient with a given disease. The UMLS Metathesaurus (Bodenreider, 2004) has over 4.4 million concepts, and the SNOMED CT ontology has 361,588 concepts (SNOMED CT, 2021). Smaller ontologies are available for phenotyping patients with human diseases, such as the Human Phenotype Ontology with 19,618 concepts (Robinson & Mundlos, 2010) and the Mendelian Inheritance of Disease in Man ontology with 94,261 concepts (McKusick, 2007). We have developed a small ontology for encoding the signs and symptoms of the neurological examination (Hier & Brint, 2020). Ideally, a disease ontology has one entry for each unique disease feature. If *brisk ankle reflex* and *brisk Achilles reflex* are exact synonyms, they are consolidated into a single concept with a unique machine-readable code.

Table 2: Phenotype of Parkinson disease in general compared to phenotype of a specific patient

Sign or Symptom	CUI Code	Disease Phenotype	This Patient Phenotype
resting tremor	C0750937	1	1
bradykinesia	C0151889	1	1
cogwheel rigidity	C0339662	1	1
stooped posture	C0344232	1	1
micrographia	C0422917	1	0
festinating gait	C0231686	1	1
expressionless face	C0234632	1	0
freezing of gait	C0860515	1	0
good response to levodopa	C1846868	1	1

1 indicates the feature is present
0 indicates the feature is absent

What phenotype repositories are available?

Three large phenotype repositories are freely available online and allow the detailed characterization of human diseases by their phenotypic features. The Human Phenotype Ontology (HPO) has over 10,000 human disease phenotype profiles (Groza et al., 2015), including rare and common diseases. The Online Mendelian Inheritance in Man (OMIM) repository has phenotype profiles for over 9,500 primarily heritable diseases (McKusick, 2007). Orphanet is a repository of genotypic and phenotypic data on 9346 rare diseases (Maiella et al., 2013). All three repositories are searchable by phenotype, gene, or disease, and each supports correlations between phenotype and genotype.

What is high throughput phenotyping?

High throughput methods allow hundreds or thousands of patient samples to be rapidly processed by automated methods. High throughput methods are available for genomics, proteomics, and metabolomics. These

methods allow large numbers of patients to be characterized quickly according to their genetic, protein, and metabolic profiles. The clinical phenotyping of individual patients remains primarily a manual process of extracting signs and symptoms (Fu et al., 2020). Several software tools are emerging to assist in this process, including *PhenoTips* (Girdea et al., 2013), *Phenolyzer* (Yang, Robinson, & Wang, 2015), and *Phenomizer* (K¨ohler et al., 2009). These software tools partially automate taking clinical text and converting it to phenotype concepts from a suitable ontology such as the Human Phenotype Ontology (HPO). Coding experts do phenotyping manually (the collection of clinical features and their conversion into machine readable codes from a disease ontology). Even among expert coders, agreement on code phenotypic features is not perfect (Chiang et al., 2006; Andrews, Richesson, & Krischer, 2007). Houle, Govindaraju, & Omholt (2010) emphasize that "gathering phenomic data is currently expensive and time-consuming; technical advances can increase phenomic throughput and lower costs."

What is a disease entity?

Finding a concise and satisfying definition of a disease entity is surprisingly difficult (Aronowitz, 2001; Magid, 2001; Hucklenbroich, 2014). As Magid (2001) has noted, "the concept of disease is inherently ambiguous and elusive." Traditionally disease entities have been defined as a complex of signs and symptoms that have diagnostic, prognostic, and therapeutic utility (Aronowitz, 2001; Hucklenbroich, 2014). The current trend is to define disease entities by their underlying genetic, physiologic, and metabolic causes. Yet, most clinicians continue to identify disease entities based on recognizable complexes of signs and symptoms. For example, clinicians continue to recognize Alzheimer disease as progressive dementia with memory loss and impaired insight, myasthenia gravis as fatigable muscle weakness with eyelid ptosis and diplopia, Huntington disease as a heritable disease with chorea, incoordination, and personality change, etc. Increasingly, consensus conferences are being used to build precise working definitions of disease entities such as Alzheimer disease, frontotemporal dementia, and Lewy body dementia (Khachaturian, 2011; McKeith et al., 2017; McKhann et al., 2001).

What is the difference between a disease entity and a disease variant?

Many diseases, and some neurological diseases, have recognizable variants that differ from the typical disease phenotype. For example, the typical amyotrophic lateral sclerosis phenotype is hyperreflexia, muscle weakness, muscle atrophy, spasticity, extensor plantar responses, and fasciculations (Grad, Rouleau, Ravits, & Cashman, 2017). A primary lateral sclerosis variant that lacks atrophy and fasciculations is recognized, as is a progressive muscle atrophy variant that lacks spasticity and hyperreflexia

(Grad et al., 2017). Similarly, the typical frontotemporal dementia phenotype includes behavioral and aphasic features. A behavioral variant of FTD, which lacks aphasic features, and an aphasic variant that lacks behavioral features is recognized (Laforce Jr, 2013). Although myasthenia gravis is characterized by fatigable generalized weakness, about 20% of patients have variant ocular myasthenia with weakness confined to the muscles of the eyes and eyelids (Sommer, Melms, Weller, & Dichgans, 1993).

What are disease mimics and disease chameleons?

Both disease mimics and disease chameleons pose diagnostic challenges for clinicians. A disease mimic is a disease that has a phenotype similar to another disease and which may be confused diagnostically with the disease diagnosis under consideration. For example, disease mimics for Parkinson disease include diseases with a similar phenotype such as Lewy body dementia, progressive supranuclear palsy, essential tremor, multiple system atrophy, dystonic tremor and others (Ali et al., 2015). Similarly, disease mimics for amyotrophic lateral sclerosis include hereditary spastic paraparesis, motor predominant chronic inflammatory demyelinating polyradiculopathy, neuralgic amyotrophy, primary progressive multiple sclerosis and others. Disease chameleons are atypical presentations of a disease with a non-standard phenotype such as Parkinson disease presenting as a pain syndrome or as a REM sleep disorder. Amyotrophic lateral sclerosis may present atypically as a flail arm or flail leg syndrome which challenges the acumen of the diagnostician (Turner & Talbot, 2013).

How can a phenotype be made computable?

Detailed descriptions of a disease phenotype are easily accessible in medical textbooks. For example, Parkinson's disease is characterized by the triad of bradykinesia, cogwheel rigidity, and resting tremor (Table 2). If disease signs and symptoms are mapped to concepts in a disease ontology (see Table 3), and each feature is binary coded (0 for absent and 1 for present), patients or diseases can be converted into computable vectors (Hier et al., 2020; Hier & Brint, 2020; Wunsch & Hier, 2021). For example, in Table 3, Parkinson disease is the vector [0,0,1,0,0,0,0,1,0,1,1] and Alzheimer disease is the vector [1,0,0,1,0,0,1,0,0,0,0].

Table 3: Example Disease Vectors based on Disease Phenotype

Finding disease	Parkinson disease	Lewy body dementia	Alzheimer disease	normal pressure hydrocephalus	frontotemporal dementia
amnesia	0	1	1	1	1
aphasia	0	0	0	0	1
bradykinesia	1	1	0	1	0
confusion	0	1	1	1	0
delusions	0	1	0	0	1
disinhibition	0	0	0	0	1
disorientation	0	1	1	1	0
gait disorder	1	1	0	1	0
incontinence	0	0	0	1	0
rigidity	1	1	0	1	0
tremor	1	1	0	0	0

Each vertical column is a disease vector so that Parkinson disease is [0,0,1,0,0,0,0,1,0,1,1]
1 indicates finding is present.
0 indicates findings are absent.

How are patient distances and patient similarities calculated?

Once a patient phenotype is converted to a vector, various distance metrics are available, including Manhattan, Cosine, Minkowski, Euclidean, Jaccard, Hamming, and others (Hier et al., 2020; Choi, Cha, & Tappert, 2010). In Table 4, the disease vectors from Table 3 have been converted into inter-disease distances. Parkinson disease is closest to Lewy body dementia and most distant to frontotemporal dementia; Alzheimer disease is closest to normal pressure hydrocephalus and most distant to Parkinson disease (Table 4). Once disease distances are known, multidimensional scaling (MDS) allows disease distances to be visualized in an arbitrary two-dimensional Cartesian space (Figure 2). The concepts of patient distances and patient similarity are closely intertwined so that if patient distances and patient similarities are normalized on a scale of 0 to 1.0, the patient distance is just the complement of the patient similarity (Hier et al., 2020; Parimbelli, Marini, Sacchi, & Bellazzi, 2018), i.e., a patient distance of 0.8 is a patient similarity of 0.2.

Table 4: Example Patient Distances based on vectors from Table 3.

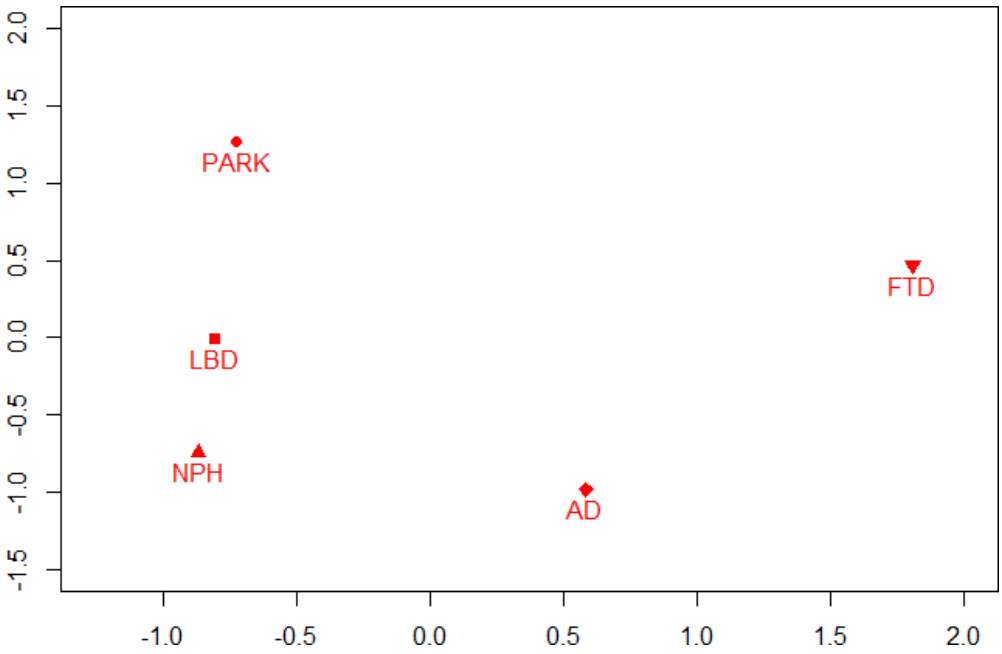
Disease	PD	LBD	AD	NH	FTD
Parkinson disease (PD)	0.00	2.00	2.65	2.24	2.83
Lewy body dementia (LBD)	2.00	0.00	2.24	1.73	2.83
Alzheimer disease (AD)	2.65	2.24	0.00	2.00	2.24
normal pressure hydrocephalus (NPH)	2.24	1.73	2.00	0.00	3.00
frontotemporal dementia (FTD)	2.83	2.83	2.24	3.00	0.00

0.00 is closest, 3.00 is most distant. Distances are Cosine distances. Note the distance of disease to itself is 0.0. Inspection of this table shows which disease is closest and most distant to each of the other diseases.

What is a pathognomonic sign?

Since the 19th century, there has been significant interest in finding pathognomonic signs and symptoms that accurately predict the presence of a disease (Janeway, 1884). A pathognomonic sign or symptom is defined as "a sign or symptom that is so characteristic of a disease that it can be used to make a diagnosis" (MedicineNet, 2021). For example, the Kayser-Fleischer ring has been considered pathognomonic of Wilson disease and Lhermitte sign pathognomonic of multiple sclerosis (Liu, Cohen, Brewer, & Laibson, 2002; Khare & Seth, 2015). Most of these attempts to find pathognomonic signs in neurology have proven prone to error (Barrows & Bennett, 1972). When phenotypic features are converted to vectors (Table 3), machine learning methods such as feature selection can be used to determine which features are most diagnostic of a disease (Wunsch & Hier, 2021; Kuhn & Johnson, 2019). When large disease datasets are available, information gain (Azhagusundari et al., 2013) and other feature selections algorithms can determine which phenotypic features are most diagnostic (i.e., most highly predictive of a given diagnosis). We have developed methods for classifying patients according to phenotypic features using machine learning (Hier et al., 2020; Wunsch & Hier, 2021). Using information gain as a selection criterion, we can ascertain which phenotypic features are most diagnostic of a given neurological diagnosis (Table 5). The best phenotypic features for diagnosing Huntington disease, Parkinson disease, Alzheimer disease, and frontotemporal dementia are chorea, resting tremor, forgetfulness, and anomia. None of these phenotypic features are pathognomonic in the classic sense, yet machine learning offers insights into which phenotypic features are quantitatively the most diagnostic.

Figure 2: Multidimensional scaling of distances from Table 4 converted into 2-dimensional Cartesian space (Demsar et al., 2013). Note the proximity of Parkinson disease (PARK), normal pressure hydrocephalus (NPH), and Lewy body dementia (LBD) which all have features of a gait disorder and hypokinesia. Note the proximity of Lewy body dementia to Parkinson disease, which both show prominent tremors. With its dementia features, the frontotemporal dementia (FTD) body is closest to Alzheimer disease (AD).



What is phenomics?

Garan) coined the term phenomics to describe the systematic study of phenotypes on a genome-wide scale while integrating basic, clinical, and information sciences (Jin, 2021. Houle et al. (2010) have emphasized that phenomics "is the acquisition of high-dimensional phenotypic data on an organism-wide scale" and that the analysis of these high dimensionality datasets (feature-rich) requires new methods and approaches. They conclude that "phenomics should be recognized and pursued as an independent discipline to enable the development and adoption of high-throughput and high-dimensional phenotyping." Recently, *Springer* has initiated a new journal dedicated to phenomics (Jin, 2021).

Table 5: Information gain towards Correct Diagnosis by Phenotypic Feature

Huntington disease	Info gain	**Parkinson disease**	Info gain	**Alzheimer disease**	Info gain	**Frontotemporal dementia**	Info gain
chorea	0.23	resting tremor	0.18	forgetful	0.12	anomia	0.06
choreic gait	0.08	cogwheel rigidity	0.15	memory loss	0.11	disinhibition	0.06
irritable	0.08	masked face	0.15	getting lost	0.11	poor hygiene	0.04
paranoia	0.04	bradykinesia	0.15	poor insight	0.06	unkempt	0.04

The top four features (most diagnostic) for each diagnosis are shown based on information gain (Azhagusundari et al., 2013). Phenotypic features have been converted to vectors (See (Wunsch & Hier, 2021)). Inspection of this table shows that the two most diagnostic phenotypic features are chorea for Huntington disease (0.23) and resting tremor for Parkinson disease (0.18).

Discussion

Precision medicine offers great promise to improve patient outcomes by integrating genomics, phenomics, metabolomics, and proteomics (Collins & Varmus, 2015). Key to this effort is the precise and computable characterization of the signs, symptoms, onset, and course of large cohorts of patients (Weng, Shah, & Hripcsak, 2020). This effort has been named high-throughput deep clinical phenotyping (Robinson, 2012; Delude, 2015). Although deep phenotyping efforts began in earnest in 2006 (Robinson, 2012), and the precision medicine initiative was introduced in 2015 (Collins & Varmus, 2015), deep phenotyping is just beginning to yield tangible results. Phenotype-based searches of online databases of Mendelian disorders such as HPO (Robinson & Mundlos, 2010) and OMIM (McKusick, 2007) are not yet performed routinely in the clinical setting (Fellner et al., 2021). Deep phenotyping of patients facilitates the conversion of unstructured clinical data into computable vectors that allow the calculation of patient distance and similarity needed for precision medicine (Brown, 2016; Parimbelli et al., 2018). Electronic health records have emerged as the most important source of phenotypic features (Curcin, 2020). A major limitation to gaining insights from deep phenotyping has been the bottleneck created by the time-consuming process of manually phenotyping patient data held in electronic health records (Curcin, 2020). Further advances in deep phenotyping may depend on improved methods to do high throughput deep phenotyping on the unstructured text of large numbers of patients that are held in electronic health records (Yehia & Eng, 2019; Yu et al., 2015; Zheng et al., 2020; Chen et al., 2013; Weng et al., 2020; Fu et al., 2020; Gehrmann et al., 2018).

This review has several limitations. Firstly, our review of the literature was focused rather than systematic. We reviewed 1372 articles and selected 53 for inclusion. Secondly, we chose examples of deep phenotyping solely

from the domain of neurology. Other domains are worthy of review as well. Thirdly, our focused review was designed as an intermediate-level introduction to the most important concepts of deep phenotyping. Each topic reviewed, including phenotype ontologies, phenotype repositories, computable phenotypes, disease variants, phenotype subtypes, high throughput phenotyping, and the algorithm-guided extraction of phenotype features from electronic health records, is worthy of exploration in greater detail than was practical in this focused review.

Disclosures

The authors have no conflicts to disclose.

References:

1. Ali, K., & Morris, H. R. (2015). Parkinson's disease: chameleons and mimics. Practical neurology, 15(1), 14-25.
2. Andrews, J. E., Richesson, R. L., & Krischer, J. (2007). Variation of SNOMED CT coding of clinical research concepts among coding experts. Journal of the American Medical Informatics Association, 14 (4), 497–506.
3. Aronowitz, R. A. (2001). When do symptoms become a disease? Annals of Internal Medicine, 134 (9), 803–808.
4. Azhagusundari, B., Thanamani, A. S., et al. (2013). Feature selection based on information gain. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2 (2), 18–21.
5. Barrows, H. S., & Bennett, K. (1972). The neurologist's diagnostic (problem-solving) skill: experimental studies and their implications for neurological training. Archives of Neurology, 26, 273-277.
6. Bodenreider, O. (2004). The unified medical language system (UMLS): integrating biomedical terminology. Nucleic Acids Research, 32 (suppl 1), D267–D270.
7. Brown, S.-A. (2016). Patient similarity: emerging concepts in systems and precision medicine. Frontiers in physiology, 7, 561.
8. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., . . . others (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature, 562 (7726), 203–209.
9. Caplan, L. (2021). Caplan-Fisher rules. Stroke, 52 (5), e155–e159.
10. Caplan, L. R. (1982). Fisher's rules. Arch Neurol, 39 (7), 389–390.
11. Chen, Y., Carroll, R. J., Hinz, E. R. M., Shah, A., Eyler, A. E., Denny, J. C., & Xu, H. (2013). Applying active learning to high-throughput phenotyping algorithms for electronic health records data. Journal of the American Medical Informatics Association, 20 (e2), e253–e259.
12. Chiang, M. F., Hwang, J. C., Alexander, C. Y., Casper, D. S., Cimino,

J. J., & Starren, J. (2006). Reliability of SNOMED-CT coding by three physicians using two terminology browsers. In Amia annual symposium proceedings (Vol. 2006, p. 131).
13. Choi, S.-S., Cha, S.-H., & Tappert, C. C. (2010). A survey of binary similarity and distance measures. Journal of systemics, cybernetics, and informatics, 8 (1), 43–48.
14. Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. New England Journal of Medicine, 372 (9), 793–795.
15. Curcin, V. (2020). Why does human phenomics matter today? Learning Health Systems, 4 (4).
16. Delude, C. M. (2015). Deep phenotyping: the details of the disease. Nature, 527 (7576), S14–S15.
17. Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., & Zupan, B. (2013). Orange: data mining toolbox in Python. The Journal of Machine Learning research, 14(1), 2349-2353. Retrieved from http://jmlr.org/papers/v14/demsar13a.html
18. Fellner, A., Ruhrman-Shahar, N., Orenstein, N., Lidzbarsky, G., Shuldiner, A. R.,
19. Gonzaga-Jauregui, C., Basel-Salmon, L. (2021). The role of phenotype-based search approaches using public online databases in diagnostics of mendelian disorders. Genetics in Medicine, 23 (6), 1095–1100.
20. Fu, S., Chen, D., He, H., Liu, S., Moon, S., Peterson, K. J., and others (2020). Clinical concept extraction: a methodology review. Journal of Biomedical Informatics, 103526.
21. Gehrmann, S., Dernoncourt, F., Li, Y., Carlson, E. T., Wu, J. T., Welt, J. (2018). Comparing deep learning and concept extraction-based methods for patient phenotyping from clinical narratives. PloS one, 13 (2), e0192360.
22. Girdea, M., Dumitriu, S., Fiume, M., Bowdin, S., Boycott, K. M., Chenier, S., et al. (2013). Phenotips: Patient phenotyping software for clinical and research use. Human Mutation, 34 (8), 1057–1065.
23. Grad, L. I., Rouleau, G. A., Ravits, J., & Cashman, N. R. (2017). Clinical spectrum of amyotrophic lateral sclerosis (ALS). Cold Spring Harbor Perspectives in Medicine, 7 (8), a024117.
24. Groza, T., Kohler, S., Moldenhauer, D., Vasilevsky, N., Baynam, G., Zemojtel, T., . . . others (2015). The human phenotype ontology: semantic unification of common and rare disease. The American Journal of Human Genetics, 97 (1), 111–124.
25. Gruber, T. (2018). Ontology. https://queksiewkhoon.tripod.com/ontology01.pdf.

26. Hier, D. B., & Brint, S. U. (2020). A neuro-ontology for the neurological examination. BMC Medical Informatics and Decision Making, 20(1), 1–9.
27. Hier, D. B., Kopel, J., Brint, S. U., Wunsch, D. C., Olbricht, G. R., Azizi, S., & Allen, B. (2020). Evaluation of standard and semantically-augmented distance metrics for neurology patients. BMC Medical Informatics and Decision Making, 20(1), 1–15.
28. Houle, D., Govindaraju, D. R., & Omholt, S. (2010). Phenomics: the next challenge. Nature Reviews Genetics, 11(12), 855–866.
29. Hucklenbroich, P. (2014). "Disease entity" as the key theoretical concept of medicine. In The Journal of medicine and philosophy: A forum for bioethics and philosophy of medicine (Vol. 39, p. 609-633).
30. Janeway, E. G. (1884, 08). Limitations of Pathognomonic signs and symptoms. Journal of the American Medical Association, III(5), 116-120. Retrieved from https://doi.org/10.1001/jama.1884.02390540004001a DOI: 10.1001/jama.1884.02390540004001a
31. Jepsen, T. C. (2009). Just what is an ontology, anyway? IT Professional Magazine, 11(5), 22.
32. Jin, L. (2021). Welcome to the Phenomics journal. Phenomics, 1, 1-2. Retrieved from https://doi.org/10.1007/s43657-020-00009-4
33. Khachaturian, Z. S. (2011). Revised criteria for diagnosis of Alzheimer's disease: National Institute on Aging-Alzheimer's Association diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 3(7), 253-256.
34. Khare, S., & Seth, D. (2015). Lhermitte's sign: the current status. Annals of Indian Academy of Neurology, 18(2), 154.
35. Kuhn, M., & Johnson, K. (2019). Feature engineering and selection: A practical approach for predictive models. CRC Press.
36. Kohler, S., Schulz, M. H., Krawitz, P., Bauer, S., D¨olken, S., Ott, C. E., et al. (2009). Clinical diagnostics in human genetics with semantic similarity searches in ontologies. The American Journal of Human Genetics, 85(4), 457–464. Retrieved from https://dx.doi.org/10.1016/j.ajhg.2009.09.003 DOI: 10.1016/j.ajhg.2009.09.003
37. Laforce Jr, R. (2013). Behavioral and language variants of frontotemporal dementia: a review of key symptoms. Clinical Neurology and Neurosurgery, 115(12), 2405–2410.
38. Liu, M., Cohen, E. J., Brewer, G. J., & Laibson, P. R. (2002). Kayser-Fleischer ring as the presenting sign of Wilson disease. American Journal of Ophthalmology, 133(6), 832–834.

39. Magid, C. S. (2001). Developing tolerance for ambiguity. JAMA, 285(1), 88–88.
40. Maiella, S., Rath, A., Angin, C., Mousson, F., & Kremp, O. (2013). Orphanet and its consortium: where to find expert-validated information on rare diseases. Revue neurologique, 169, S3-8.
41. McKeith, I. G., Boeve, B. F., Dickson, D. W., Halliday, G., Taylor, J.-P., Weintraub, D., et al. (2017). Diagnosis and management of dementia with Lewy bodies: Fourth consensus report of the DLB consortium. Neurology, 89(1), 88–100.
42. McKhann, G. M., Albert, M. S., Grossman, M., Miller, B., Dickson, D., & Trojanowski, J. Q. (2001). Clinical and pathological diagnosis of frontotemporal dementia: report of the workgroup on frontotemporal dementia and pick's disease. Archives of neurology, 58(11), 1803–1809.
43. McKusick, V. A. (2007). Mendelian Inheritance in Man and its online version, OMIM. American Journal of Human Genetics, 80(4), 588.
44. MedicineNet. (2021). Pathognomonic. https://www.medicinenet.com/pathognomonic/definition.htm.
45. Parimbelli, E., Marini, S., Sacchi, L., & Bellazzi, R. (2018). Patient similarity for precision medicine: A systematic review. Journal of biomedical informatics, 83, 87–96.
46. Robinson, P. N. (2012). Deep phenotyping for precision medicine. Human Mutation, 33(5), 777–780.
47. Robinson, P. N., & Mundlos, S. (2010). The human phenotype ontology. Clinical Genetics, 77(6), 525–534.
48. SNOMED CT. (2021). NCBO BioPortal. https://bioportal.bioontology.org/ontologies/SNOMEDCT.
49. Sommer, N., Melms, A., Weller, M., & Dichgans, J. (1993). Ocular myasthenia gravis. Documenta Ophthalmologica, 84(4), 309–333.
50. Tasker, R. C. (2017). Why everyone should care about "computable phenotypes." Pediatric critical care medicine: a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies, 18(5), 489.
51. Turner, M. R., & Talbot, K. (2013). Mimics and chameleons in motor neurone disease. Practical neurology, 13(3), 153-164.
52. Weng, C., Shah, N. H., & Hripcsak, G. (2020). Deep phenotyping: embracing complexity and temporality—towards scalability, portability, and interoperability. Journal of biomedical informatics, 105, 103433.
53. Wunsch, D. C., & Hier, D. B. (2021). Subsumption reduces dataset dimensionality without decreasing the performance of a machine learning classifier. 43rd Annual International Conference of the IEEE

Engineering in Medicine and Biology Society (EMBC), 2021, 1622-1625.

54. Yang, H., Robinson, P. N., & Wang, K. (2015). Phenolyzer: phenotype-based prioritization of candidate genes for human diseases. Nature Methods, 12(9), 841–843.
55. Yehia, L., & Eng, C. (2019). Largescale population genomics versus deep phenotyping: brute force or elegant pragmatism towards precision medicine. Nature Publishing Group.
56. Yu, S., Liao, K. P., Shaw, S. Y., Gainer, V. S., Churchill, S. E., Szolovits, P., Cai, T. (2015). Toward high-throughput phenotyping: unbiased automated feature extraction and selection from knowledge sources. Journal of the American Medical Informatics Association, 22(5), 993–1000.
57. Zheng, N. S., Feng, Q., Kerchberger, V. E., Zhao, J., Edwards, T. L., Cox, N. J., Wei, W.-Q. (2020). Phemap: a multi-resource knowledge base for high-throughput phenotyping within electronic health records. Journal of the American Medical Informatics Association, 27(11), 1675–1687.