Resource Summary Report

Generated by T1D on Sep 19, 2026

PheKB

RRID:SCR_005292
Type: Tool

Proper Citation

PheKB (RRID:SCR_005292)

Resource Information

URL: http://phenotype.mc.vanderbilt.edu/

Proper Citation: PheKB (RRID:SCR_005292)

Description: Collaborative environment of building and validating electronic phenotype algorithms using electronic medical records (EMRs) and natural language processing (NLP) for use in genome-wide association studies (GWAS). On this site you can: View existing algorithms, Enter or create new algorithms, Collaborate with others to create or review algorithms, View implementation details for existing algorithms. The Electronic Medical Records and Genomics Network (eMERGE) has investigated whether data captured through routine clinical care using electronic medical records (EMRs) can identify disease phenotypes with sufficient positive and negative predictive values for use in genome-wide association studies (GWAS). Most EMRs captured key information (diagnoses, medications, laboratory tests) used to define phenotypes in a structured format; in addition, natural language processing has also been shown to improve case identification rates. PheKB is an outgrowth of that validation effort. Phenotype algorithms can be viewed by data modalities or methods used: CPT codes, ICD 10 codes, ICD 9 codes, Laboratories, Medications, Vital Signs, Natural Language Processing Algorithms can also be viewed by: * Implementation results (positive predictive value, sensitivity, publications) * Institution * Work Group

Abbreviations: PheKB

Synonyms: Phenotype KnowledgeBase, PheKB - a knowledgebase for discovering phenotypes from electronic medical records

Resource Type: knowledge environment, software repository, software resource

Defining Citation: PMID:20362271

Keywords: phenotype, electronic medical record, medical record, human, clinical, white blood cell, red blood cell, lipid, algorithm, height, cardiac conduction, genome-wide association study, natural language processing

Related Condition: Atrial fibrillation, Crohn""s disease, Multiple Sclerosis, Rheumatoid arthritis, Type 2 diabetes mellitus, Dementia, Cataracts, Hypothyroidism, Diabetic Retinopathy, High-Density Lipoprotein, Peripheral Arterial Disease

Resource Name: PheKB

Resource ID: SCR_005292

Alternate IDs: nlx_144339

Record Creation Time: 20220129T080229+0000

Record Last Update: 20260919T195051+0000

Ratings and Alerts

No rating or validation information has been found for PheKB.

No alerts have been found for PheKB.

Data and Source Information

Source: SciCrunch Registry

Usage and Citation Metrics

We found 32 mentions in open access literature.

Listed below are recent publications. The full list is available at T1D .

Pan H, et al. (2026) Replication of COPD susceptibility loci in a large Chinese elderly population using a validated, multi-setting EHR phenotype. BMC pulmonary medicine, 26(1), 47.

Lapinska S, et al. (2026) Evaluation of the genome-informed risk assessment (GIRA) approach from eMERGE in an independent health system. medRxiv : the preprint server for health sciences.

O'Rorke MA, et al. (2026) Assessing Patient Characteristics in Neuroendocrine Tumor Research: A Comparison of the NET-PRO Study to SEER Population-Based Data. Endocrine, 91(1).

Xian S, et al. (2026) GWAS analysis of a depression cohort defined by an EHR-phenotyping algorithm reveals the role of immune regulations in depression risk. Frontiers in genetics, 17, 1818653.

Binkheder S, et al. (2025) Biomedical literature-based clinical phenotype definition discovery using large language models. Database : the journal of biological databases and curation, 2025.

Rattsev I, et al. (2025) A Framework to Quantify Disparities in Pharmacogenomic Treatment Concordance and Drug Response Outcomes. medRxiv : the preprint server for health sciences.

Fishe J, et al. (2025) COMPAC: COMputable Phenotype for Asthma in Children. Research square.

Johnson R, et al. (2024) Unified Clinical Vocabulary Embeddings for Advancing Precision Medicine. medRxiv : the preprint server for health sciences.

Xu J, et al. (2024) Combining Federated Machine Learning and Qualitative Methods to Investigate Novel Pediatric Asthma Subtypes: Protocol for a Mixed Methods Study. JMIR research protocols, 13, e57981.

Abend AH, et al. (2024) Estimation of prevalence of autoimmune diseases in the United States using electronic health record data. The Journal of clinical investigation, 135(4).

Ritchie SC, et al. (2024) Integrated clinical risk prediction of type 2 diabetes with a multifactorial polygenic risk score. medRxiv : the preprint server for health sciences.

Yu J, et al. (2022) Under-specification as the source of ambiguity and vagueness in narrative phenotype algorithm definitions. BMC medical informatics and decision making, 22(1), 23.

Berchuck SI, et al. (2022) A Framework for Automating Psychiatric Distress Screening in Ophthalmology Clinics Using an EHR-Derived AI Algorithm. Translational vision science & technology, 11(10), 6.

Binkheder S, et al. (2022) PhenoDEF: a corpus for annotating sentences with information of phenotype definitions in biomedical literature. Journal of biomedical semantics, 13(1), 17.

Almowil ZA, et al. (2021) Concept libraries for automatic electronic health record based phenotyping: A review. International journal of population data science, 6(1), 1362.

Chapman M, et al. (2021) Desiderata for the development of next-generation electronic health record phenotype libraries. GigaScience, 10(9).

May SB, et al. (2021) A Phenotyping Algorithm to Identify People With HIV in Electronic Health Record Data (HIV-Phen): Development and Evaluation Study. JMIR formative research, 5(11), e28620.

Sivasankar S, et al. (2021) Use of large scale EHR data to evaluate A1c utilization among sickle cell disease patients. BMC medical informatics and decision making, 21(1), 268.

Lee J, et al. (2021) Comparative effectiveness of medical concept embedding for feature engineering in phenotyping. JAMIA open, 4(2), ooab028.

De Freitas JK, et al. (2021) Phe2vec: Automated disease phenotyping based on unsupervised embeddings from electronic health records. Patterns (New York, N.Y.), 2(9), 100337.