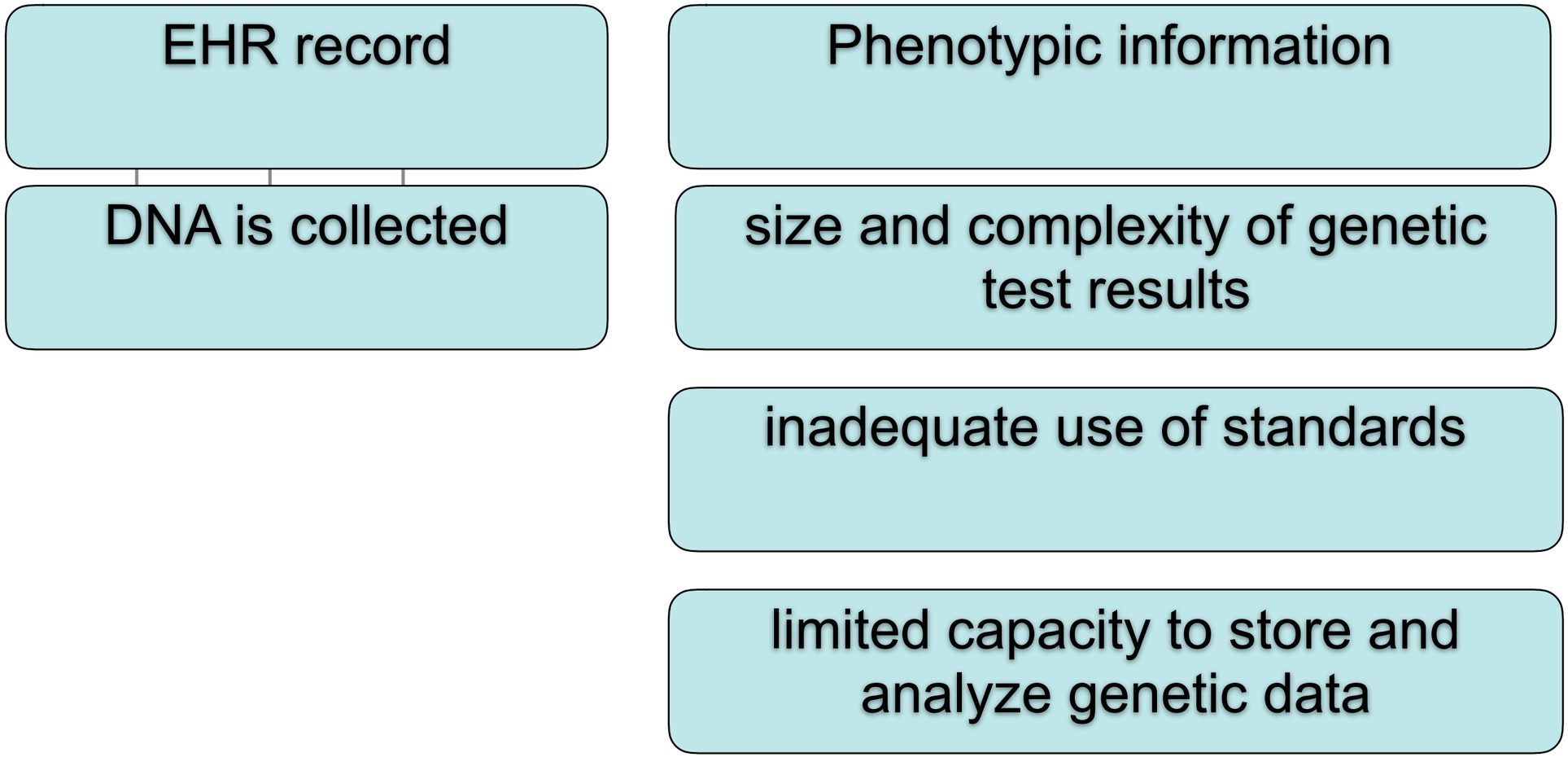
Genome Integration with electronic health records/ clinical information

MEDT32/33

- Return actionable results from variant data to electronic health records (EHRs)
- Incorporation of genetics into the EHR for clinical decision support


EHR record
DNA is collected
Phenotypic information
size and complexity of genetic test results
inadequate use of standards
limited capacity to store and analyze genetic data

- EHRs are designed primarily for clinical care
- data availability
- missing data
- incorrect data
- unstructured narrative text

Wei and Denny *Genome Medicine* (2015) 7:41
DOI 10.1186/s13073-015-0166-y

Genome Medicine

REVIEW Open Access

Extracting research-quality phenotypes from electronic health records to support precision medicine

Wei-Qi Wei[1] and Joshua C Denny[1,2*]

Abstract

The convergence of two rapidly developing technologies - high-throughput genotyping and electronic health records (EHRs) - gives scientists an unprecedented opportunity to utilize routine healthcare data to accelerate genomic discovery. Institutions and healthcare systems have been building EHR-linked DNA biobanks to enable such a vision. However, the precise extraction of detailed disease and drug-response phenotype information hidden in EHRs is not an easy task. EHR-based studies have successfully replicated known associations, made new discoveries for diseases and drug response traits, rapidly contributed cases and

arthritis) and observable traits (for example, height, skin pigmentation or drug response). Similarly, more recent efforts to look at rare variants through next-generation sequencing technologies have identified causative SNPs for rare diseases [3] as well as important modulators for some common diseases [4-6]. Through these efforts, genetic determinants of many human diseases and, more recently, therapeutic responses, are being deciphered.

Traditionally, genetic studies have leveraged purpose-built cohorts [7,8] (such as the Wellcome Trust Consortium [9], Framingham Heart Study [10] and Human Heredity and Health in Africa Consortium [11]). These studies often use self-report questionnaires and/or clinical staff to obtain participant phenotypes. While this approach provides qual-

Structured information

- billing codes
- laboratory reports
- variables
  - physiologic measurements
  - demographic information

Key info

- clinical notes
- test results
  - Pathology
  - Radiology
- combinations:
  - Free-text search + billing codes + manual chart review.

Scale

- large cohorts necessary for genome-wide association studies
- natural language processing methods to process narrative text data
- generate high-validity collections of cases and controls

BioBanks and EHR

- EHR-derived cases can be used for genomic discovery and validation
- enable multiple genomic investigations upon a single set of genotyped individuals
- genotyped once as part of a study on diabetes
- then participate in another analysis for cardiovascular disease
- enables phenome-wide association studies of genetic variants

NLP

- NLP efforts to extract medical concepts from clinical text documents focused on coding in the Systematic Nomenclature of Pathology
- ICD for financial and billing purposes
- Unified Medical Language System (UMLS)
- SNOMED-CT

- filter out concepts (or keywords) within a corpus of documents that indicate statements *other than the patient having the disease*
- phenotype classification include identifying family medical history context and negated terms (e.g., "*no cardiac disease*")

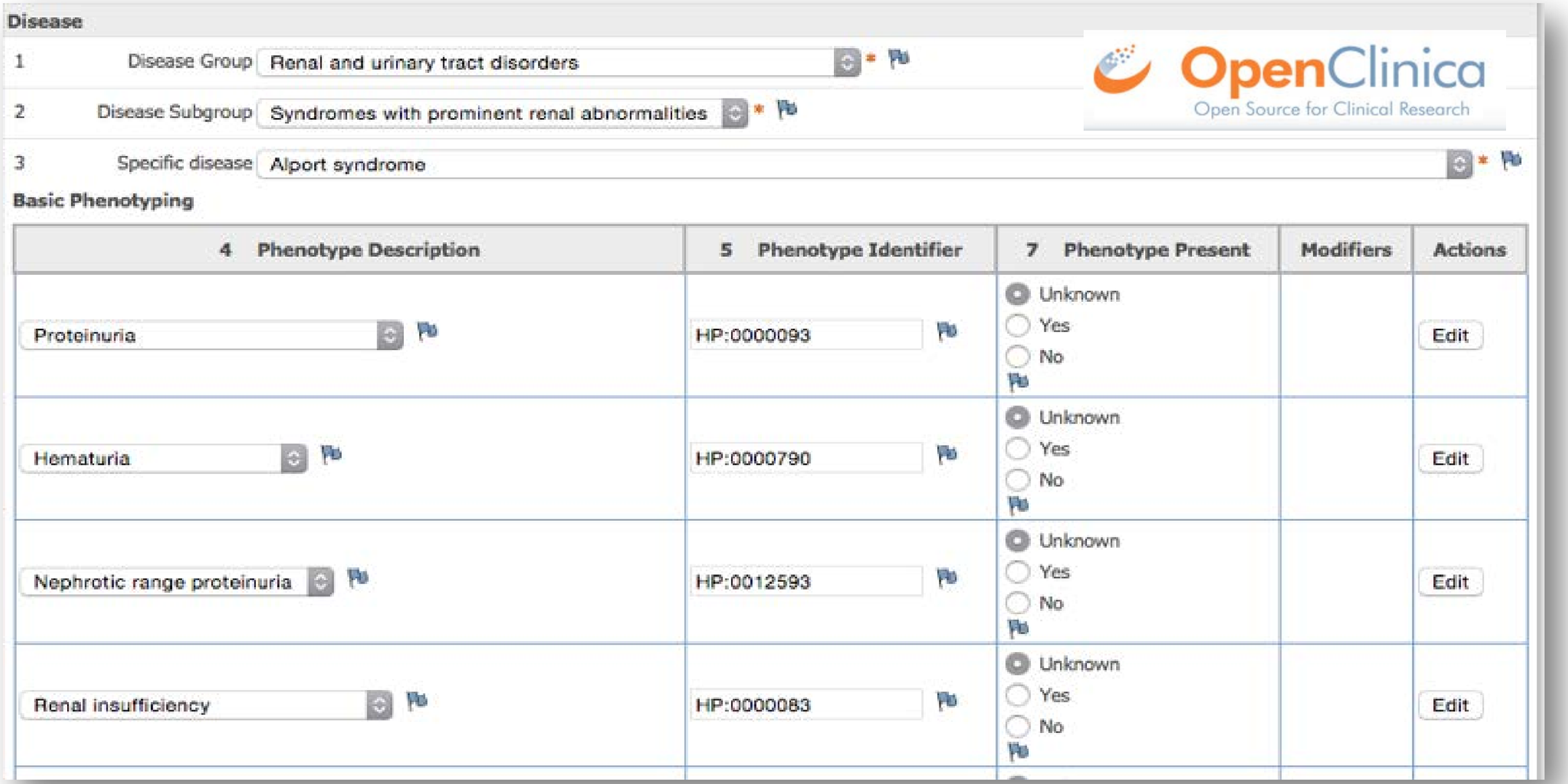
National Joint Registry (NJR)

- framework for large-scale genomic discovery using routinely collected national clinical audit data
- distance recruitment of 40,000 individuals with knee injuries
- identify epidemiological risk factors
- genes and pathways that are associated with different patterns of knee OA
- disease progression and poor clinical outcomes

How do genes predict response to drugs?

- https://www.youtube.com/watch?v=icNMEqoDCAg
- minute 2 onwards dominate 29 - PHARMGKB

OpenClinica phenotype entry



Disease

1 Disease Group: Renal and urinary tract disorders *

2 Disease Subgroup: Syndromes with prominent renal abnormalities *

3 Specific disease: Alport syndrome *

OpenClinica
Open Source for Clinical Research

Basic Phenotyping

4 Phenotype Description	5 Phenotype Identifier	7 Phenotype Present	Modifiers	Actions
Proteinuria	HP:0000093	Unknown / Yes / No		Edit
Hematuria	HP:0000790	Unknown / Yes / No		Edit
Nephrotic range proteinuria	HP:0012593	Unknown / Yes / No		Edit
Renal insufficiency	HP:0000083	Unknown / Yes / No		Edit

Additional terms not present in the data model can be naturally added

Genomics england

relate gene variants to pharmacokinetic impacts