**COMP 353/453 Project Phase 1**

Deliverables:

* Description of Organization
* ER Diagram including min/max pairs and candidate keys
* Constraints outside ER diagram
* Relational schema

Assessment:

* Status report
* Phase summary

Phase Leader Crisi Patelis \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Phase Recorder Michael Saban \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Phase Checker Sathvik Maridasana Nagaraj \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Technical Advisor Aaron Myrold \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. **Introduction**

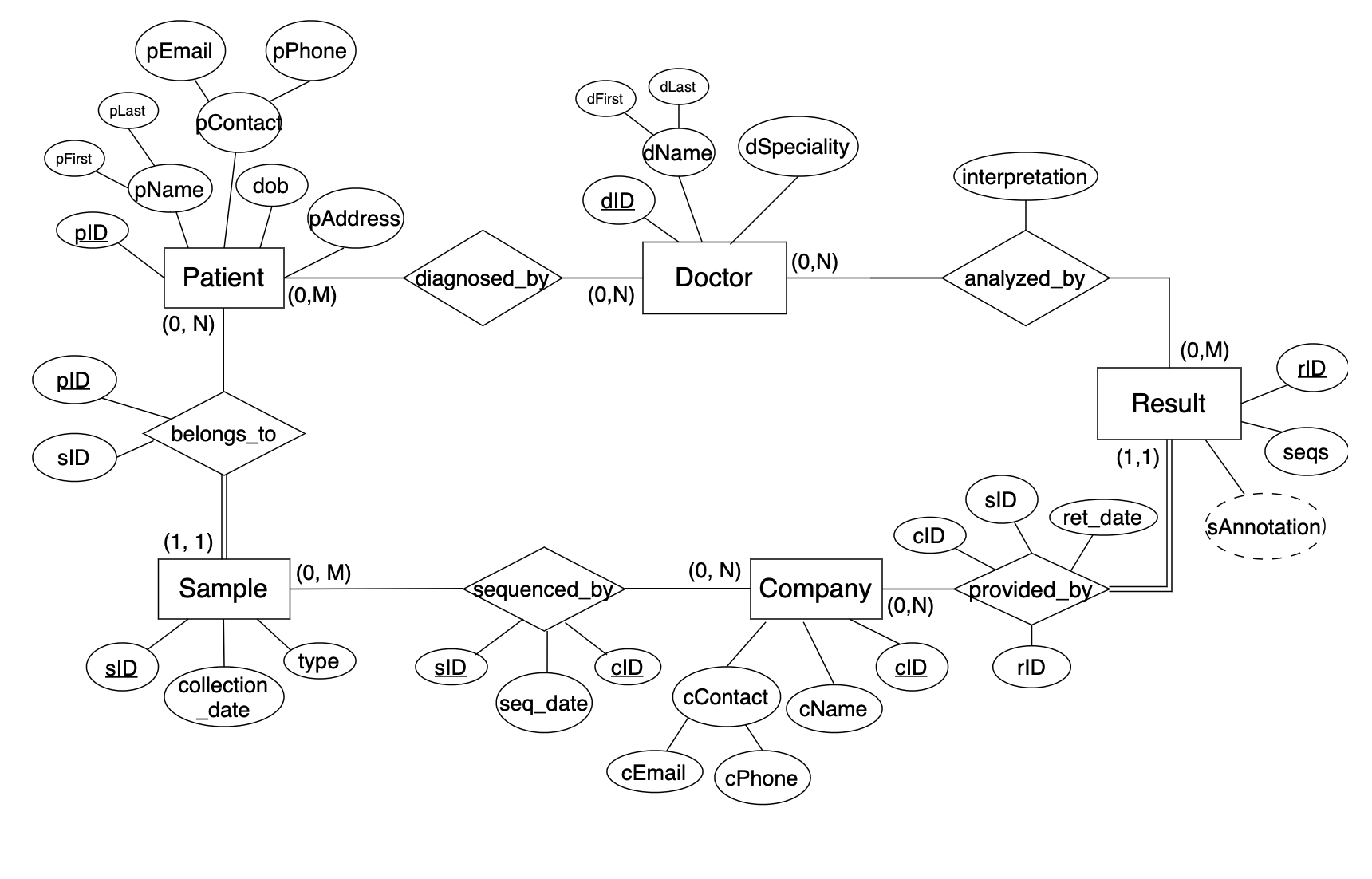
The database is based on a tool that Nick Miller has worked on called Flype. It consists of a relational database, in-house developed bioinformatics analysis pipelines, a connection framework to connect to external tools, and a web interface to provide easy and efficient access to the information stored in the database. The relational database stores things like patient information, sample and result information, result interpretation, and other information needed to provide quick and personalized healthcare.

1. **Description of the Organization**

The Healthcare database stores patient information along with their data related to test results such as Next Generation Sequencing (NGS) and Pharmacogenomic (PGX) testing. Maintaining a relational database that not only holds all patient information and results, but also interpretation of said results provides a quick and efficient way for healthcare providers to incorporate computational biology and genetic testing into a more personalized treatment for patients.

* Patients require genetic testing for a variety of illnesses. This involves collecting the correct samples and sending them off for sequencing. To efficiently track this, patients are given a unique patient ID along with storing their names, birth dates, and addresses. Patients may have multiple doctors working with them at one time or none at all.
* Over time, patients may require multiple different tests, so each sample includes a unique sample ID, a collection date, and information on the type of sample collected. Patients can have multiple samples taken, but each sample can only belong to one patient.
* A single collected sample may be sequenced by multiple companies, so contact information including the company name, email, and phone number are included along with a unique company ID. This allows tracking of communications with each company and when and where samples are sent. Every individual company may be sequencing multiple samples.
* The results returned by the company are associated with the sample and patient IDs and include not only the sequencing results, but also identified and annotated gene variants of the sequenced samples. A result can only belong to one company, but a company may send multiple results back.
* While sequencing results contain annotation, a healthcare professional needs to analyze/interpret and relay that information to the patient’s electronic health record (EHR). Doctors’ names and specialties are stored under a unique ID. Doctors can have multiple patients or none at all.

1. **ER Diagram**



1. **ER Diagram Uncaptured Constraints**

* Sequence strings in Results must be uppercase

1. **Relational Schema**

**5.1 Relational Schema with Referential Integrity**

Patient(pID,pFirst,pLast,pDOB,pAddress,pEmail,pPhone)

Sample(sID,collection\_date,type,pID)

Foreign key (pID) references Patient (pID)

Sequenced\_By(cID,sID,seq\_date)

Foreign key (cID) references Company (cID)

Foreign key (sID) references Sample (sID)

Company(cID,cName,cEmail,cPhone)

provided\_by(cID,rID, ret\_date)

Foreign key (cID) references Company (cID)

Foreign key (rID) references Result (rID)

Result(rID,seqs,sAnnotation,sID)

Foreign key (sID) references Sample (sID)

analyzed\_by(rID,dID,interpretation)

Foreign key (rID) references Result (rID)

Foreign key (dID) references Doctor (rID)

Doctor(dID,dFirst,dLast,dSpecialty)

diagnosed\_by(dID,pID,diag\_date)

Foreign key (dID) references Doctor (dID)

Foreign key (pID) references Patient (pID)

**5.2 Relational Table Details**

| **Table Name** | **Attribute** | **Description** |
| --- | --- | --- |
| Patient | pID | unique patient ID |
| pFirst | patient’s first name |
| pLast | patient’s last name |
| pDOB | patient’s date of birth |
| pAddress | patient’s address |
| pEmail | patient’s email address |
| pPhone | patient’s phone number |
| Sample | sID | unique sample ID |
| collection\_date | date of sample’s collection |
| type | type of sample |
| pID | patient ID |
| Sequenced\_by | cID | unique company ID |
| sID | unique sample ID |
| sequence\_date | Date sample was sequenced |
| Company | cID | unique company ID |
| cName | company name |
| cEmail | company email address |
| cPhone | company phone number |
| Results | rID | result ID |
| seqs | sequences |
| sAnnotation | sequence annotation |
| provided\_by | cID | Company ID |
| sID | Sequence ID |
| rID | Result ID |
| ret\_date | Date sequenced sample was returned |
| Doctor | dID | unique doctor ID |
| dFirst | doctor’s first name |
| dLast | doctor’s last name |
| dSpecialty | doctor’s specialty |
| analyzed\_by | rID | Result ID |
| dID | Doctor ID |
| interpretation | Doctor’s interpretation |
| Diagnosed\_by | dID | unique doctor ID |
| pID | unique patient ID |

**Group Status Report**

Dates and attendance of team meetings:

2/9 All

2/16 All

Progress overview as of February 16th:

Data requirements for the project database have been mostly identified. Potential to change pending communication with Nick Miller. The initial deliverable includes the description of the relational database organization, the ER diagram with corresponding min/max pairs, a list of constraints not in the ER diagram, and the relational schema for the ER diagram.

Contributions of group members:

Phase Leader: Crisi Patelis

* Organized deliverable
* Added relational table descriptions
* Translated word processed ER diagram to diagram.net
* Created patient, doctor, and sample to ER

Phase Recorder: Michael Saban

* Word processed ER diagram
* Detailed constraints not in ER diagram
* Added relational schema section
* Prepared organization description
* Flushed out patient, doctor, and sample attributes/relations on ER

Phase Checker: Sathvik Maridasana Nagaraj

* Aided in ER diagram organization and construction
* Aided in status report and phase summary
* Flushed out results attributes/relations on ER

Technical Advisor: Aaron Myrold

* Aided in ER diagram organization and construction
* Prepared phase summary
* Aided in introduction
* Added results entity and attributes to ER