WAREHOUSE

Getting Started – The User Experience

April 6, 2016

Hauwa Yusuf

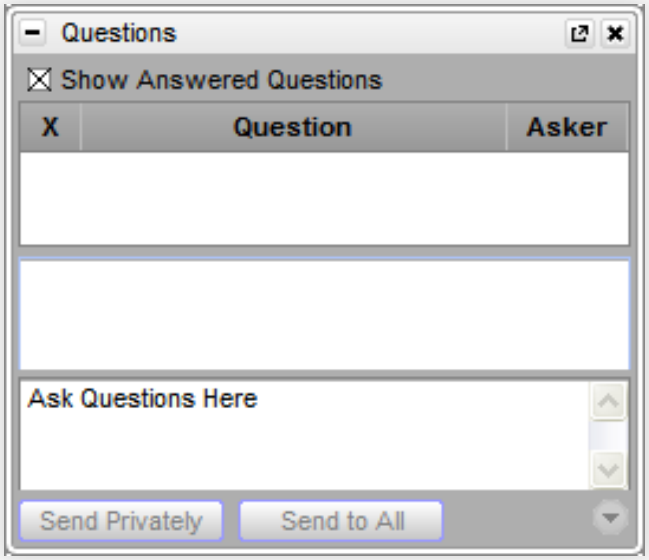
Field Application Scientist

GOLDEN HELIX

Accelerating the Quest for Significance™

Questions During The Presentation

Use the Questions pane in your GoToWebinar window



Agenda

1. Overview Golden Helix
2. Introduction to Genetic Data Warehousing
3. VSWarehouse - Concepts and Use Cases
4. Q & A

Golden Helix – Who We Are

Golden Helix is a global bioinformatics company founded in 1998.

gsk GlaxoSmithKline

SNP & VARIATION SUITE

GWAS
Genomic Prediction
Large-N-Population Studies
RNA-Seq
CNV-Analysis

varSEQ®

Filtering and Annotation
Clinical Reports
Pipeline
Data Warehousing

Over 300 customers globally

SickKids
THE HOSPITAL FOR SICK CHILDREN

Cincinnati Children's
Hospital Medical Center

The Children's Hospital of Philadelphia

Stanford University

UBC

Ucla

Yale

The Feinstein Institute for Medical Research
North Shore-Long Island Jewish Health System

North Shore LIJ

PREVENTION GENETICS
DISEASE PREVENTION THROUGH GENETIC TESTING

MAYO CLINIC

LINEAGEN

MOFFITT CANCER CENTER

Dartmouth-Hitchcock
NORRIS COTTON CANCER CENTER

BAYER

Lilly

abbvie

TEVA

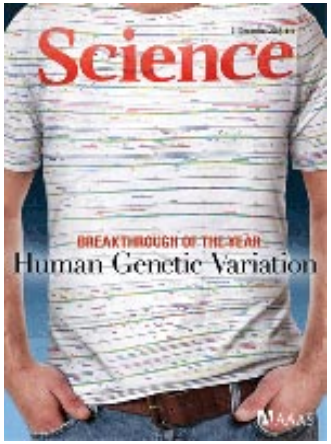
NATIONAL CANCER INSTITUTE

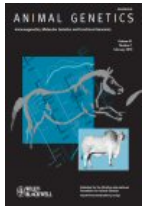
USDA
USDA Agricultural Research Service

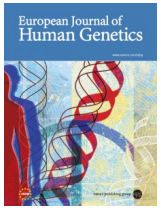
FDA

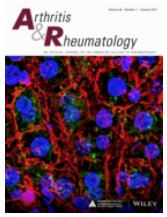
GOLDEN HELIX
Accelerating the Quest for Significance™

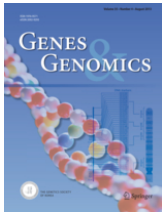
Cited in over 1000 peer-reviewed publications

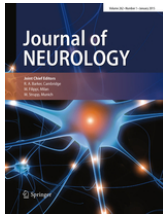








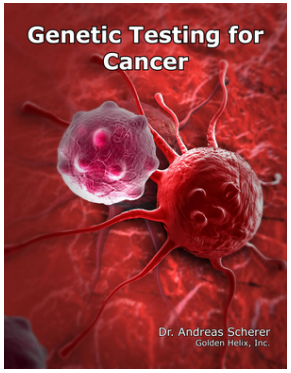




Golden Helix – Who We Are

When you choose a Golden Helix solution, you get more than just software

- REPUTATION
- TRUST
- EXPERIENCE

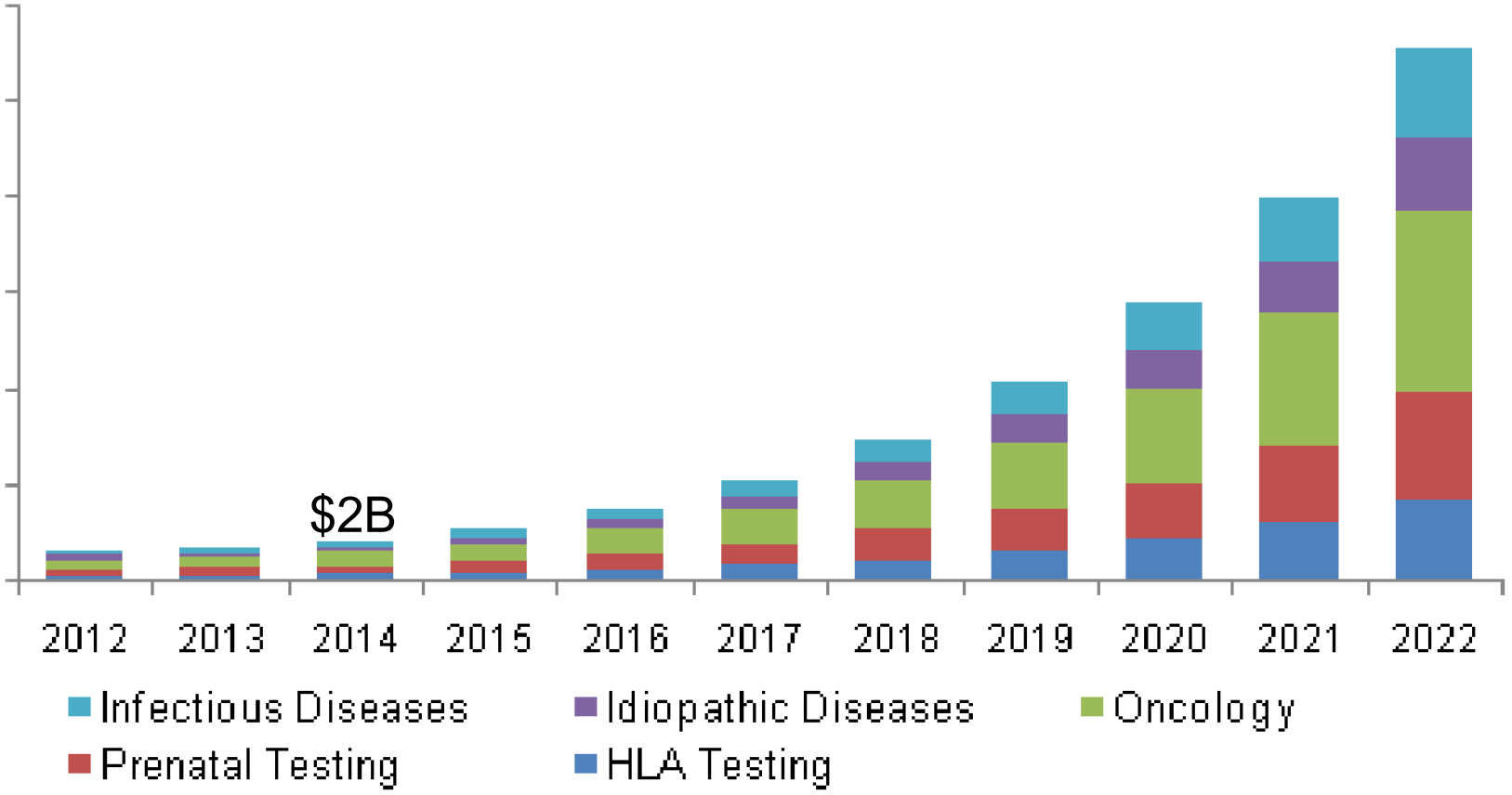


- INDUSTRY FOCUS
- THOUGHT LEADERSHIP
- COMMUNITY

- TRAINING
- SUPPORT
- RESPONSIVENESS

- TRANSPARENCY
- INNOVATION and SPEED
- CUSTOMIZATIONS

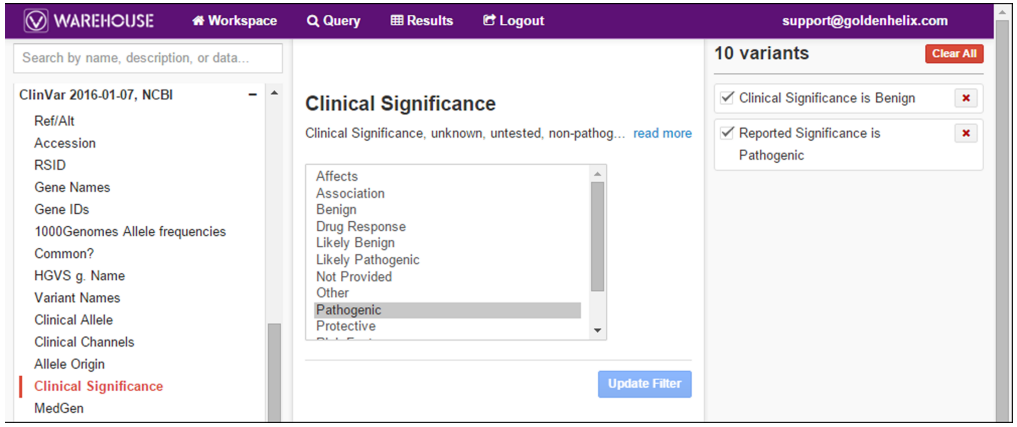
Precision Medicine unfolding

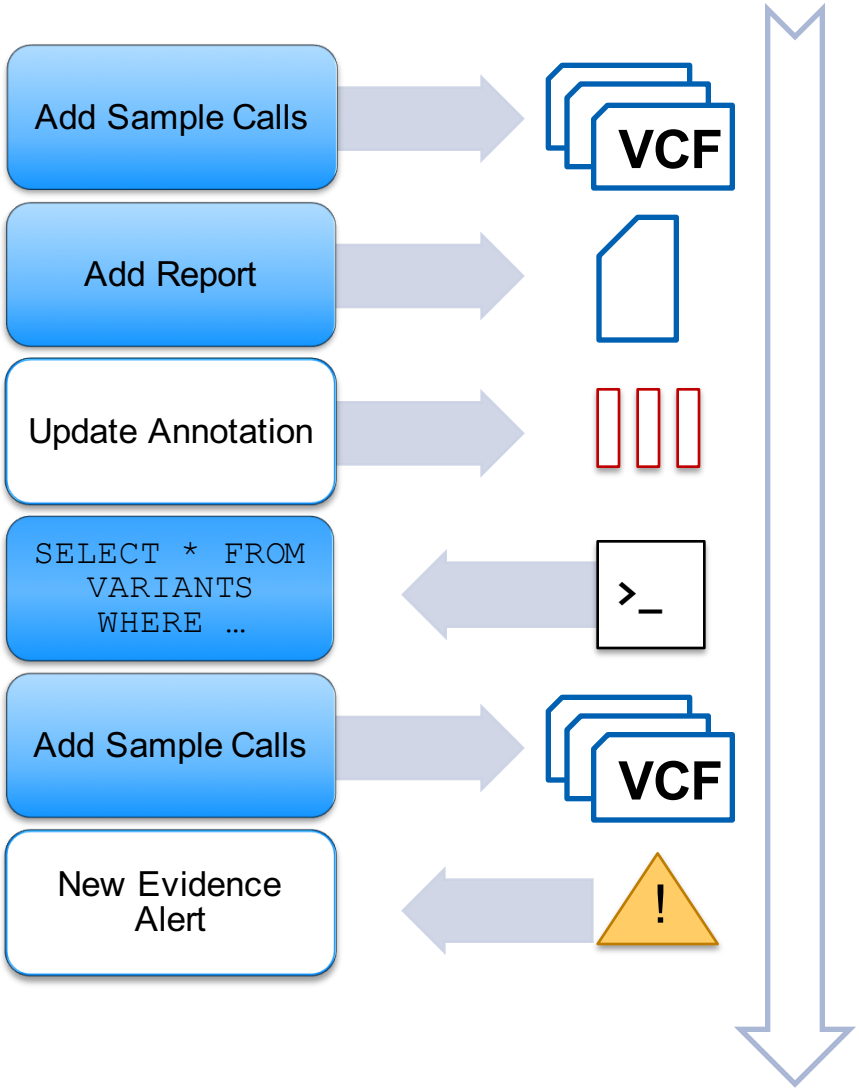


Grand View Market Research

Requirements for Variant Warehousing

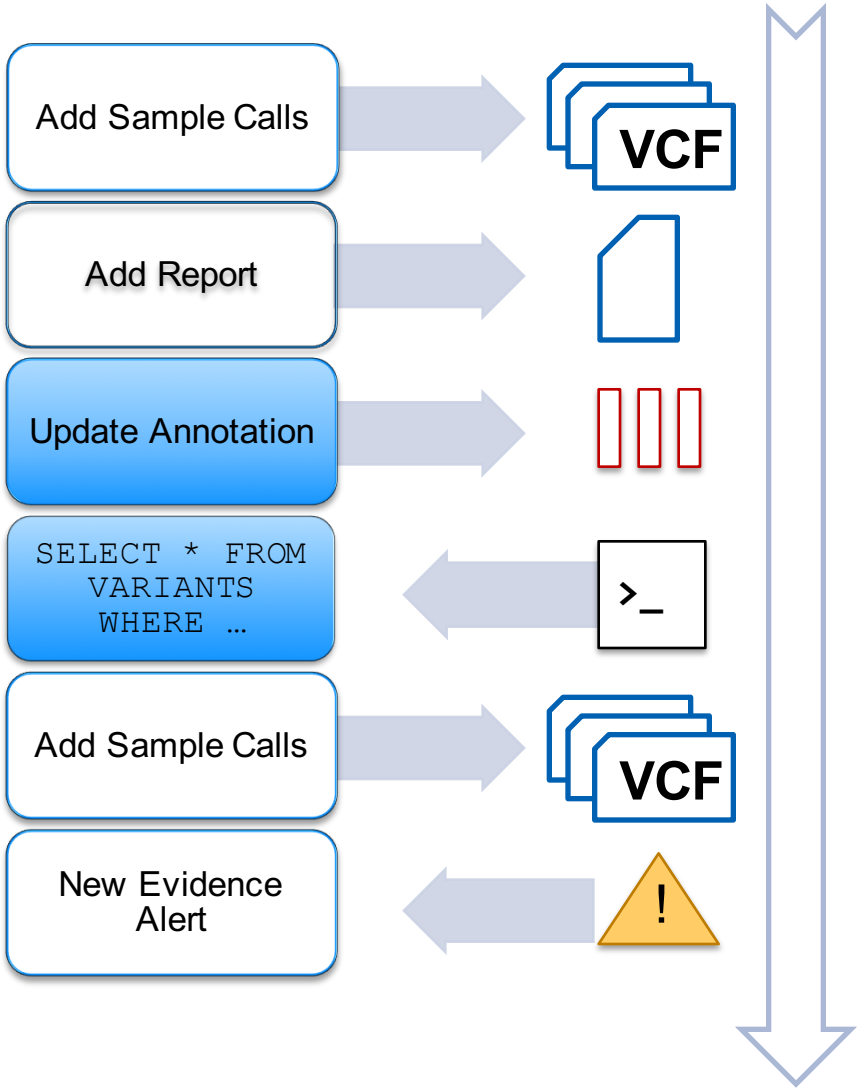
- Archive full VCFs of every sequenced sample
- Archive Reports
- Query and retrieve subsets of data at any time





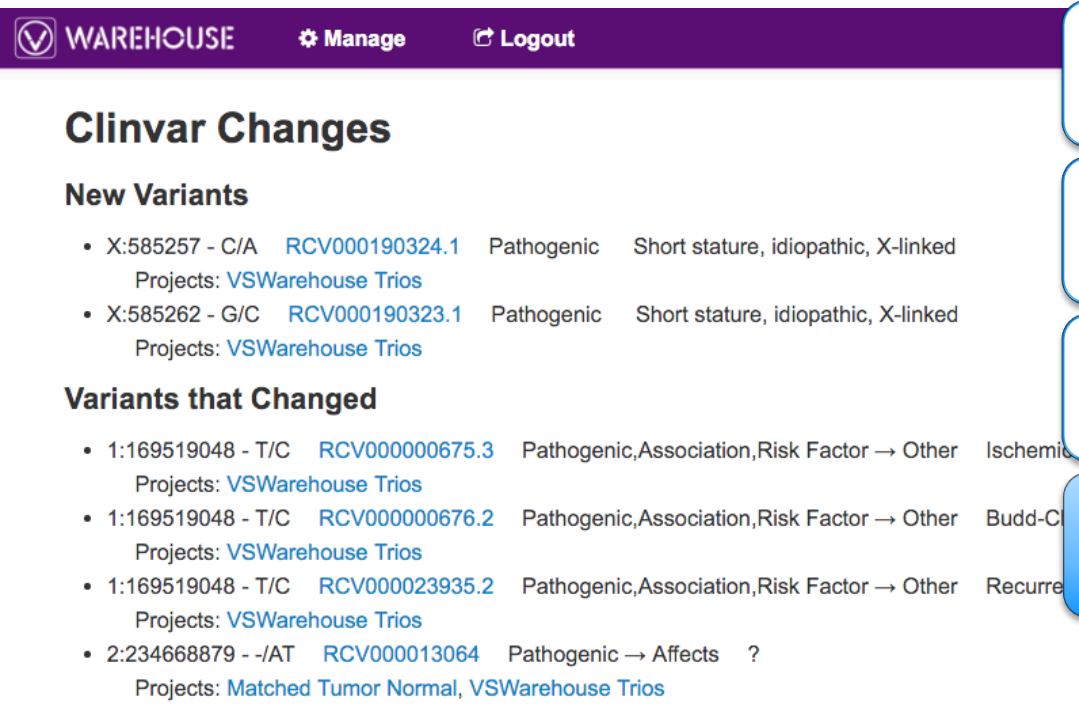
Requirements for Variant Warehousing

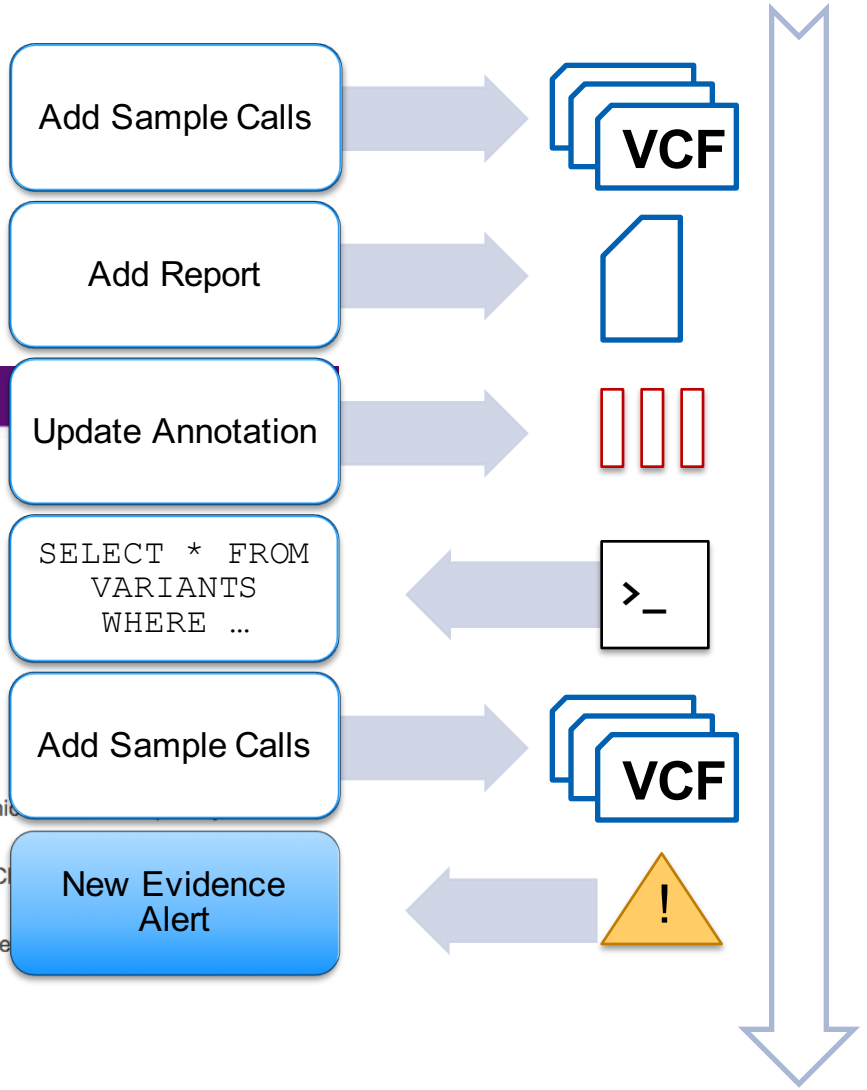
- **Annotation Source**
  - Have I ever seen this variant in my previous test samples?
  - At what frequency?
  - Does this gene contain other rare variants in my cohort?
  - Have I classified this variant?
  - Have I used this variant in a report for any previous samples?



Requirements for Variant Warehousing

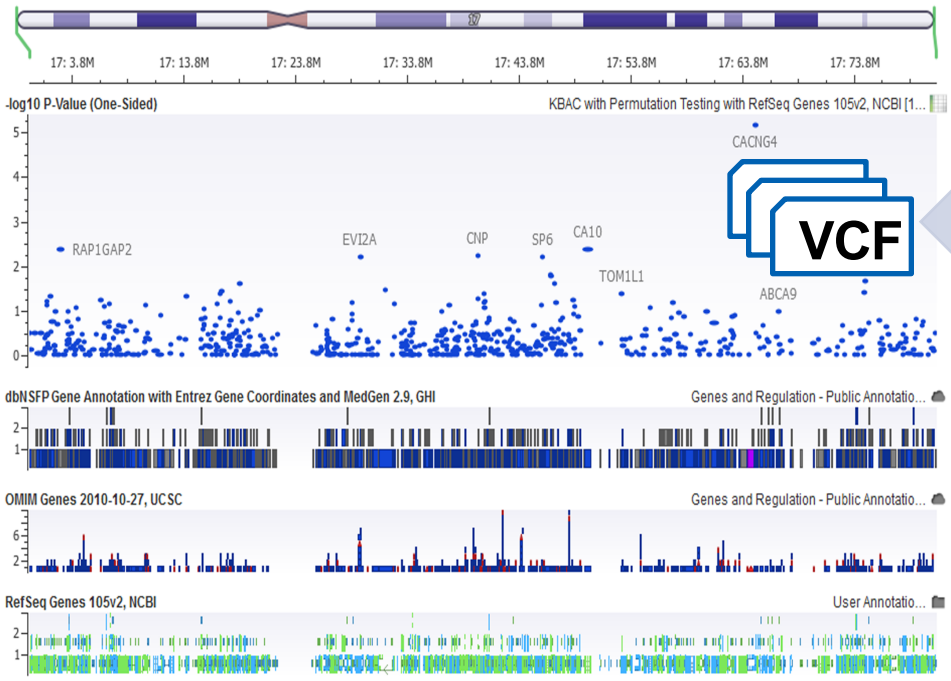
- **Clinical Alerts**
  - ClinVar's monthly release has new and updated variant classifications.
  - Has the classification of a variant changed?

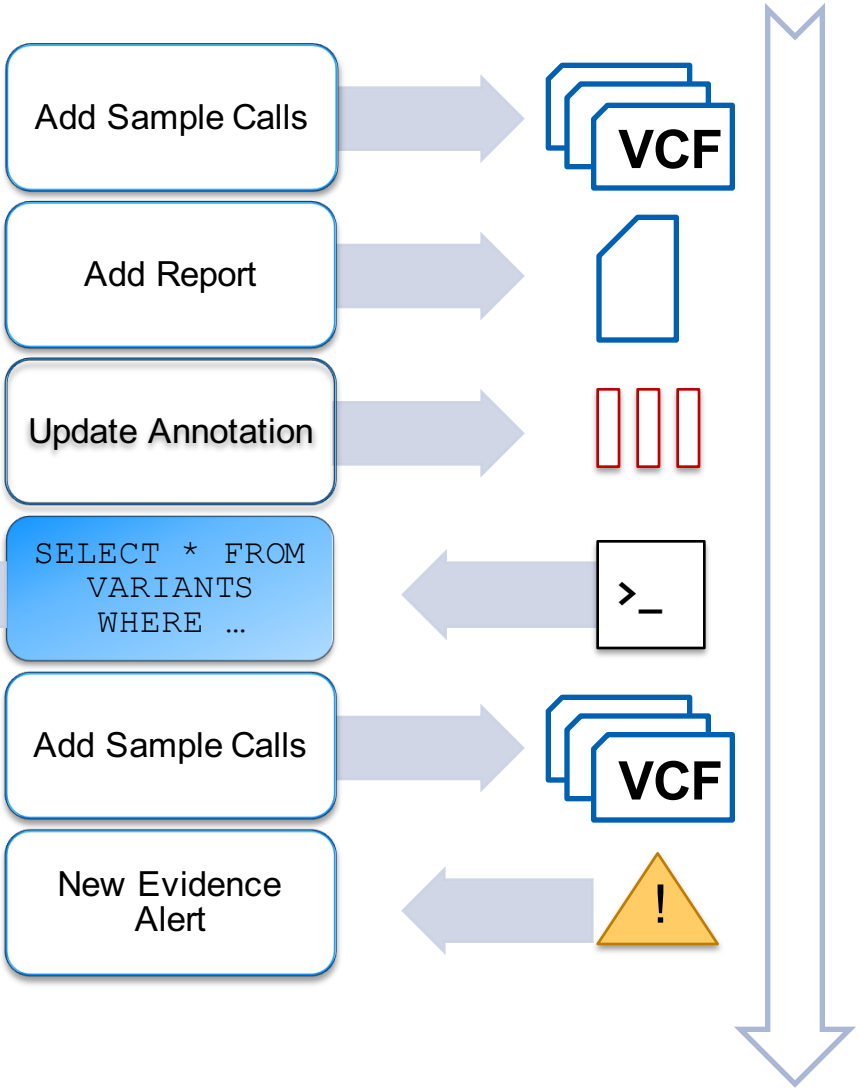




Requirements for Variant Warehousing

- **Conducting Research**
  - Capturing samples and reports
  - Extract affected and unaffected study participants to conduct further research on a genomic level.





Requirements for Variant Warehousing

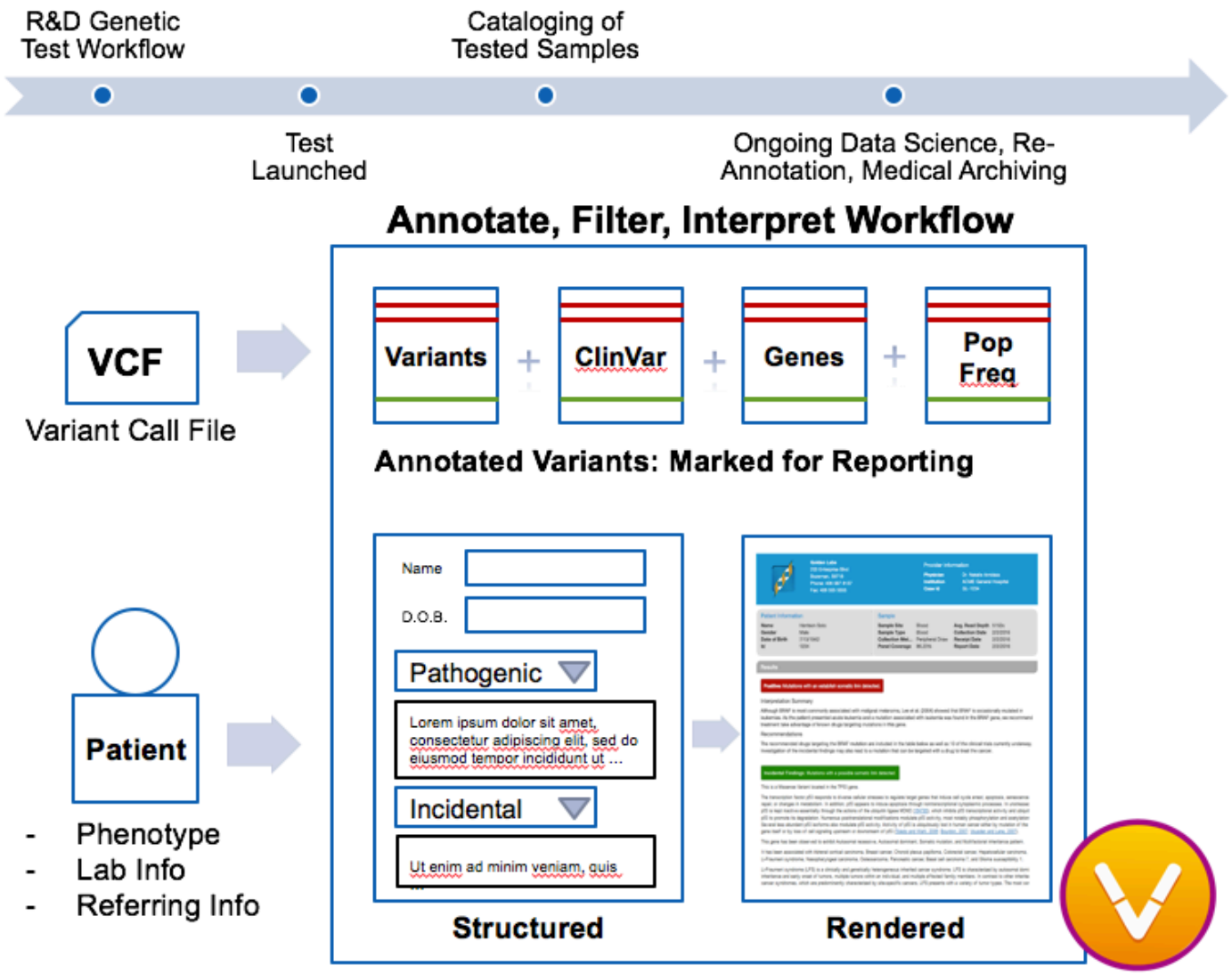
- **Connecting with other legacy systems**
  - Integration point between lab and other hospital systems.
  - How many tests did I conduct in the past day/week/month/year?
  - How many samples did we add to the warehouse in the past day/week/month/year?

```
Baldur:Desktop ghi$ cat warehousequery.py
#!/usr/bin/python
import psycopg2
from prettytable import PrettyTable

# connect to database
connection = psycopg2.connect("dbname='ghivarseq'
                              user='ghi'
                              host='warehouse.goldenhelix.
                              password='ghipassword'")
cursor = connection.cursor()

query = ("SELECT count(*) "
         "FROM cancerpanel_1 "
         "WHERE report_generated=True;")

cursor.execute(query)
results = cursor.fetchall()

output table
table = PrettyTable(["Number_of_Reports"])
for result in results:
    table.add_row(result)

Baldur:Desktop ghi$ python warehousequery.py
+-------------------+
| Number_of_Reports |
+-------------------+
|         5         |
+-------------------+
```

```
import requests
import json

fields  = [{'category':'Variant Sites', 'field':'Chromosome'},
           {'category':'Variant Sites', 'field':'Pos Start'},
           {'category':'Variant Sites', 'field':'Ref/Alt'},
           {'category':'Variant Sites', 'field':'Identifier'},
           {'category':'Summary of COSMIC Mutations Left Aligned 71 v2, GHI','field':'Mutation ID

filters = [
            {
              'category': 'RefSeq Genes 105v2, NCBI',
              'field': 'Gene Names',
              'operator': 'overlaps',
              'value': ['BRCA1', 'SAMD11', 'OR4F5', 'NOCL2']
            },
            {...
            },
            {...
            },
            {...
            }
          ]

query = {'fields':fields, 'filters':filters}

response = requests.post('http://warehouse.goldenhelix.com/api/project_variant/CEU Trio/3/query/'
                         data={'query':json.dumps(query)},
                         auth=('user@goldenhelix.com', 'password'))
```

API – SQL, REST, Python

VarSeq Workflow



[Demonstration]

WAREHOUSE

CEU Trio

Query

View Report

Clinical Alerts

Export Samples

YRI Trio

Annotate

Create Report

Update Report

Update Project

VSWarehouse Use Cases

1. Annotation Source
2. Clinical Alerts
3. Research
4. API

Questions or More Info

- Email info@goldenhelix.com
- Request an evaluation of the software at www.goldenhelix.com

Questions?

Use the Questions pane in your GoToWebinar window