Developing clinical decision support systems for cancer and COVID-19 precision medicine

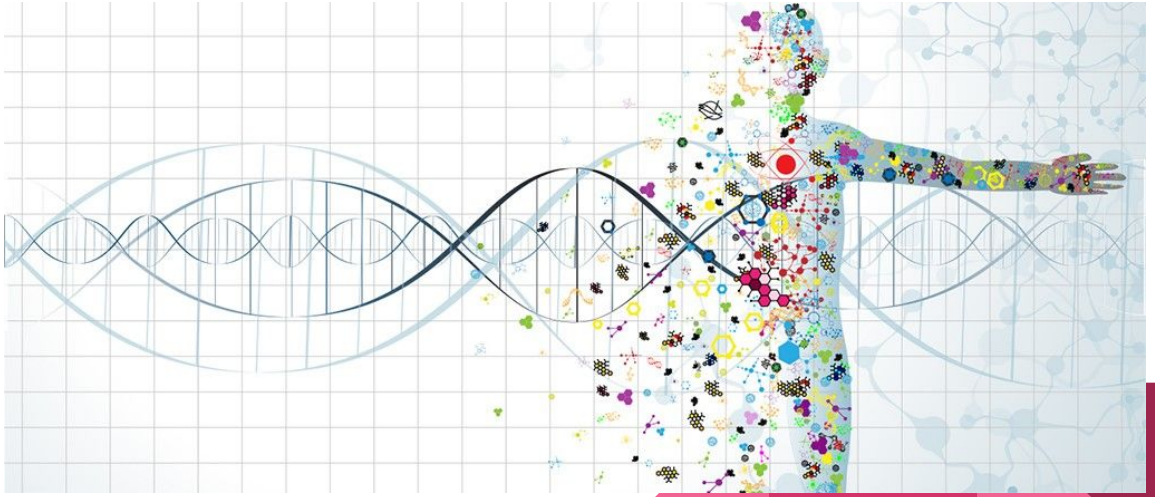
Varun Suraj
Mentor: Dr. Gil Alterovitz

Motivations

Growth in Precision Medicine

“An emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person.”

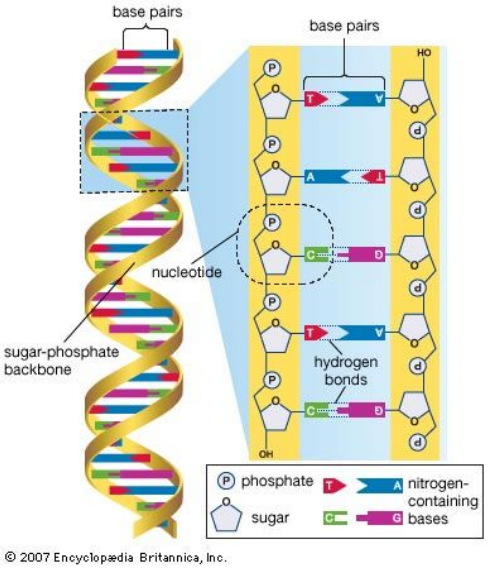
- Precision Medicine Initiative, NIH



Availability of Electronic Patient Data

“EHRs are real-time, patient-centered records that make information available instantly and securely to authorized users.”

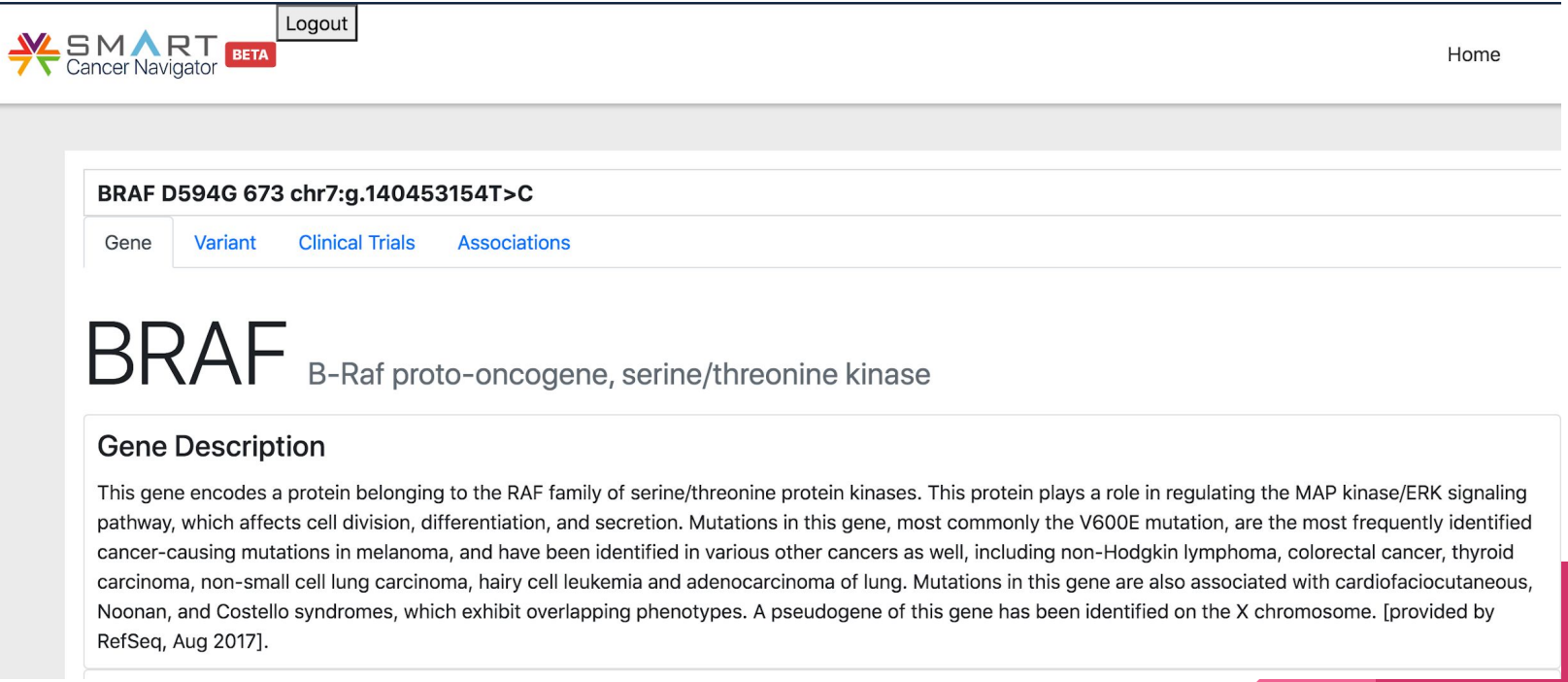
- HealthIT.gov

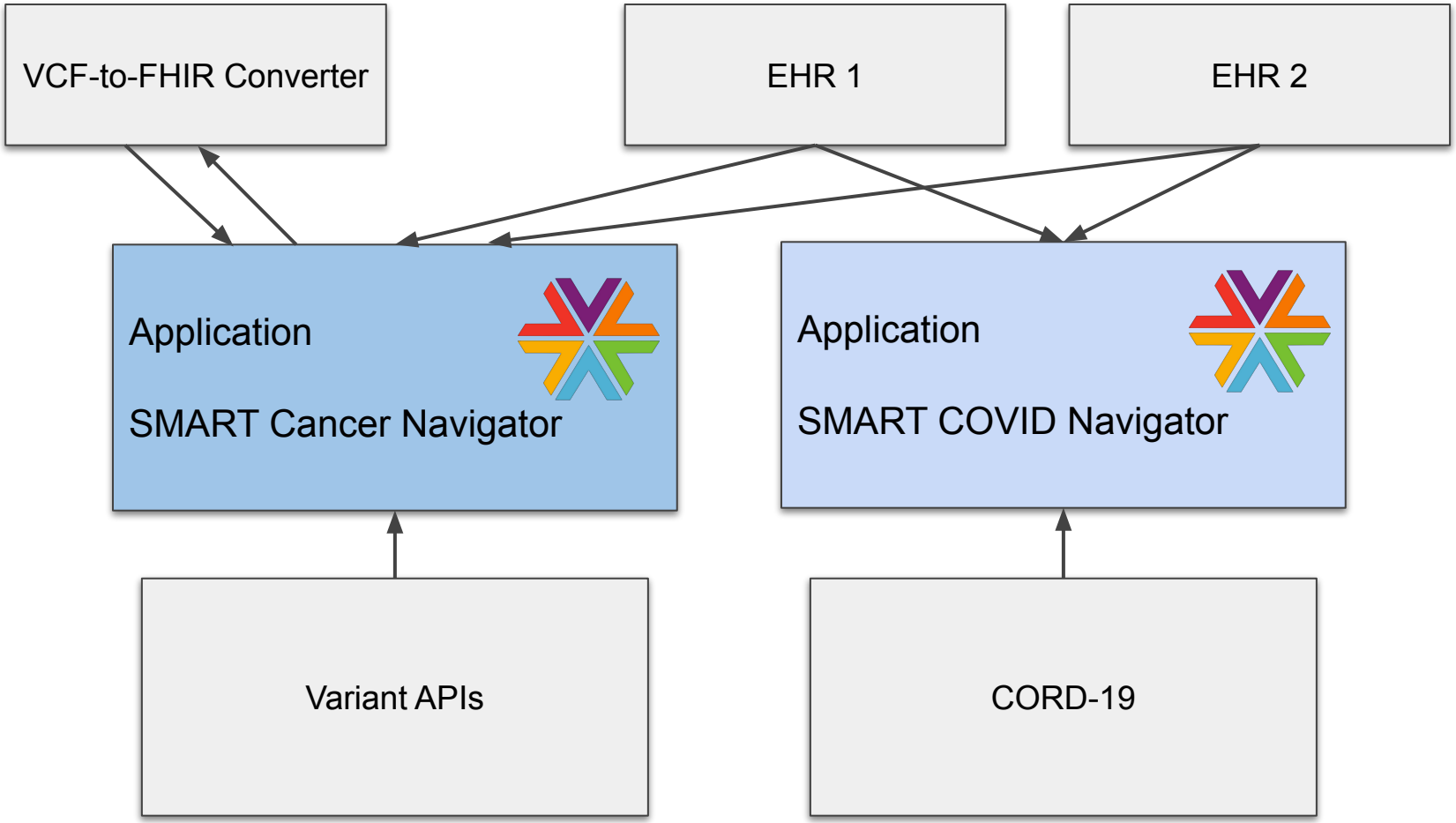


Sequencing DNA means determining the order of the four chemical building blocks - called "bases" - that make up the DNA molecule.

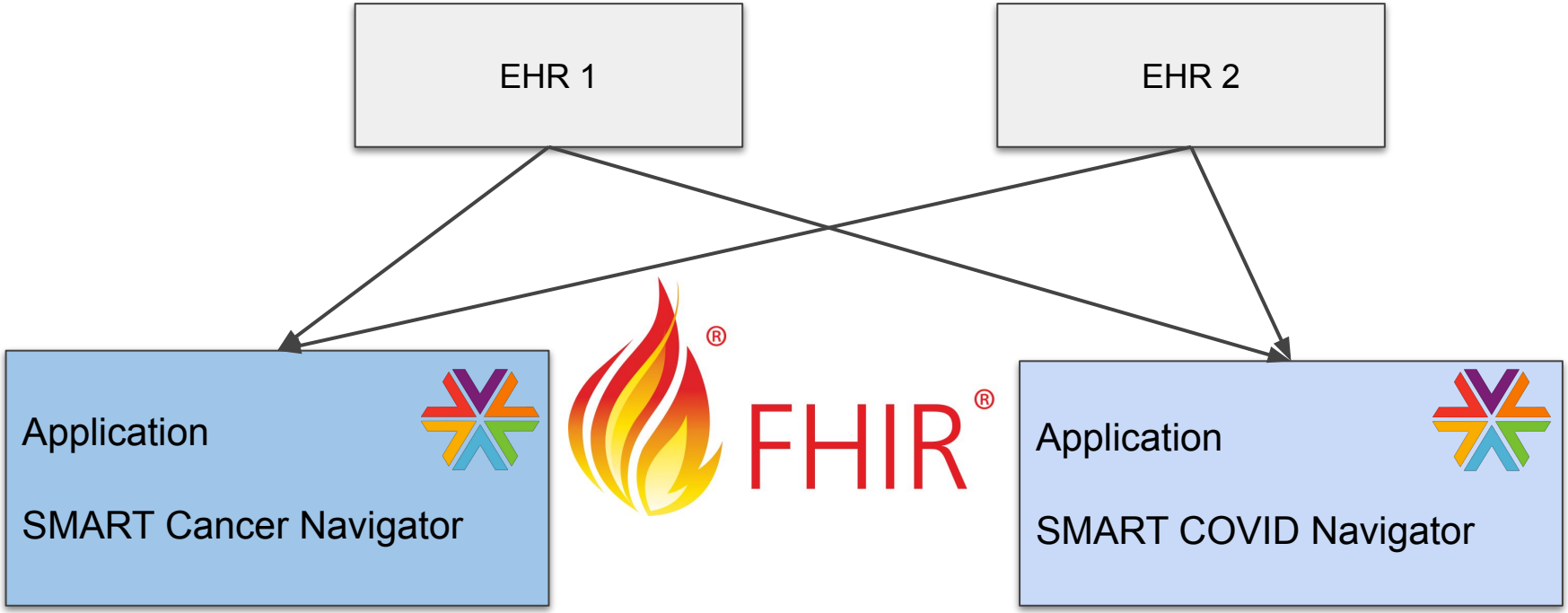
- National Human Genome Research Institute

SMART Cancer Navigator

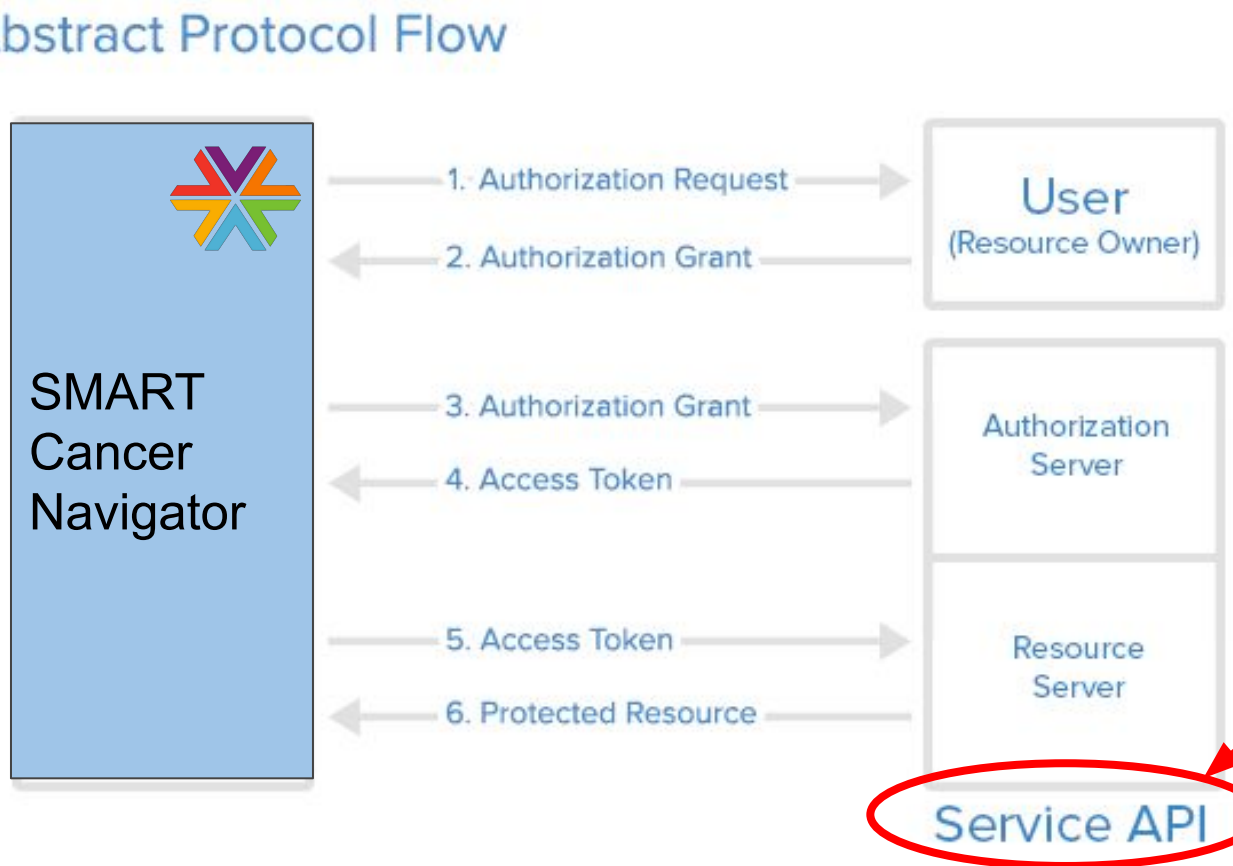


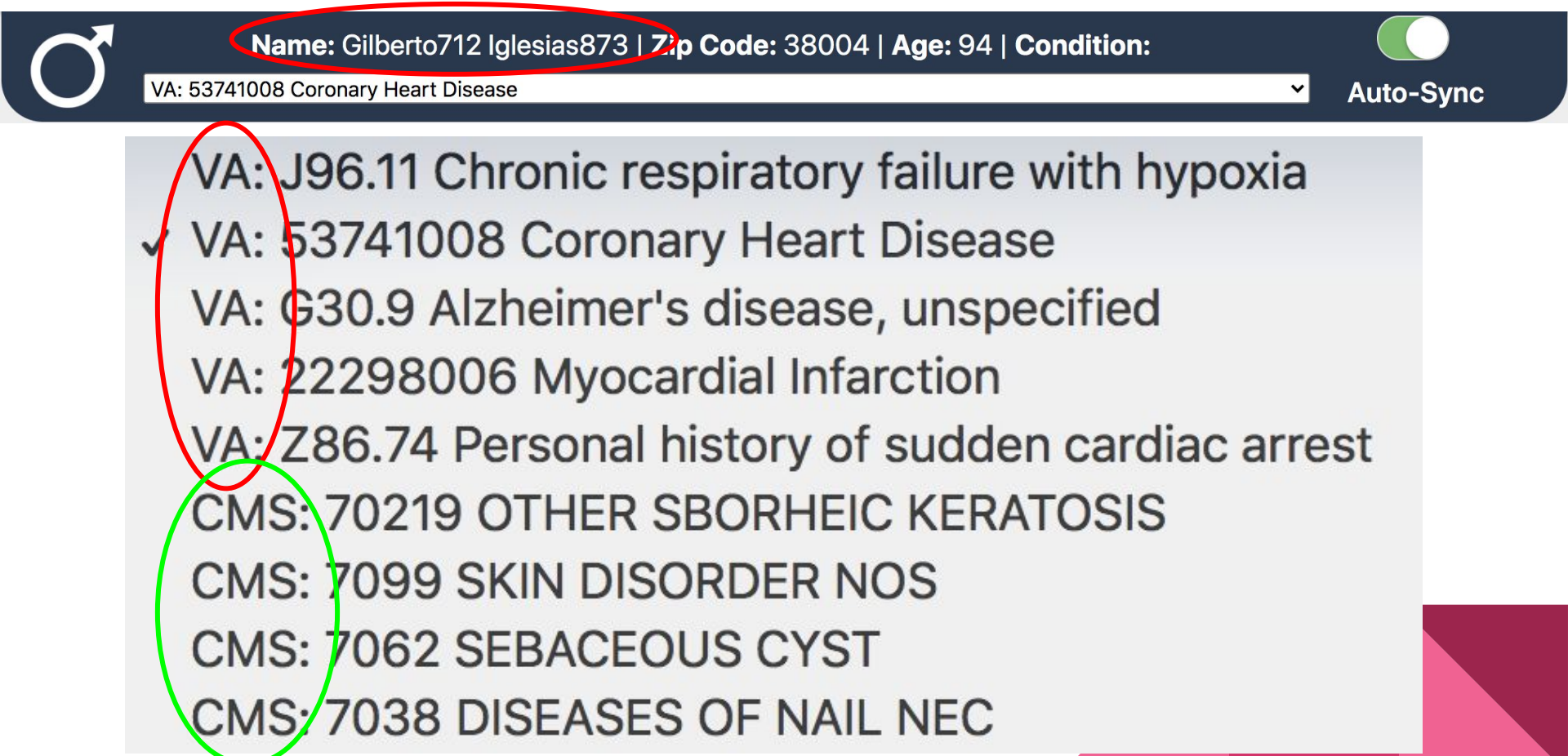

VCF-to-FHIR Converter
EHR 1
EHR 2
Application
SMART Cancer Navigator
Application
SMART COVID Navigator
Variant APIs
CORD-19

Phase 1: Multiple EHRs

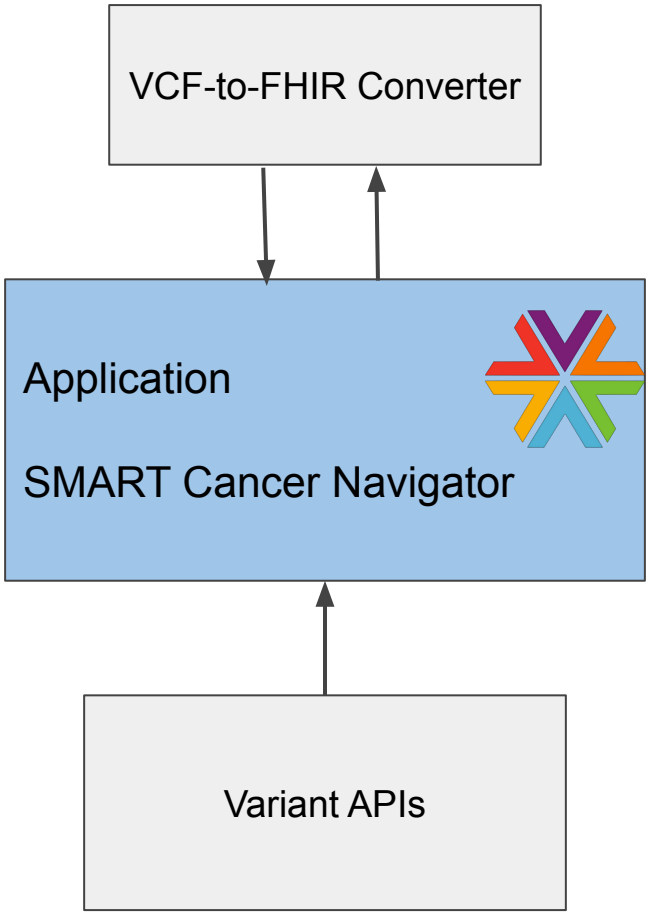


OAuth2 Authorization Process




Name: Gilberto712 Iglesias873 | Zip Code: 38004 | Age: 94 | Condition:
VA: 53741008 Coronary Heart Disease
Auto-Sync
VA: J96.11 Chronic respiratory failure with hypoxia
VA: 53741008 Coronary Heart Disease
VA: G30.9 Alzheimer's disease, unspecified
VA: 22298006 Myocardial Infarction
VA: Z86.74 Personal history of sudden cardiac arrest
CMS: 70219 OTHER SBORHEIC KERATOSIS
CMS: 7099 SKIN DISORDER NOS
CMS: 7062 SEBACEOUS CYST
CMS: 7038 DISEASES OF NAIL NEC

Phase 2: Adding Genomic Data

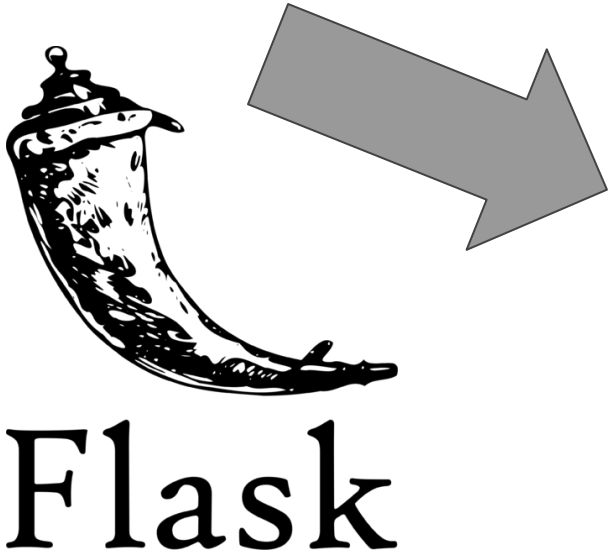


Sample VCF File

```
##fileformat=VCFv4.3
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS      ID         REF   ALT    QUAL  FILTER   INFO                              FORMAT       NA00001         NA00002          NA00003
20     14370    rs6054257  G     A      29    PASS     NS=3;DP=14;AF=0.5;DB;H2           GT:GQ:DP:HQ  0|0:48:1:51,51  1|0:48:8:51,51   1/1:43:5:.,.
20     17330    .          T     A      3     q10      NS=3;DP=11;AF=0.017               GT:GQ:DP:HQ  0|0:49:3:58,50  0|1:3:5:65,3     0/0:41:3
20     1110696  rs6040355  A     G,T    67    PASS     NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ  1|2:21:6:23,27  2|1:2:0:18,2     2/2:35:4
20     1230237  .          T     .      47    PASS     NS=3;DP=13;AA=T                   GT:GQ:DP:HQ  0|0:54:7:56,60  0|0:48:4:51,51   0/0:61:2
20     1234567  microsat1  GTC   G,GTCT 50    PASS     NS=3;DP=9;AA=G                    GT:GQ:DP     0/1:35:4        0/2:17:2         1/1:40:3
```

```
#CHROM POS     ID        REF    ALT    QUAL  FILTER INFO  FORMAT      NA12878
10     96448001  .       T      <CGA_CNVWIN> .     .      NS=1;CGA_WINEND=9
10     96448129  .       A      G      .     .      NS=1;AN=2;AC=1;CGA_XR=d
18.5   GT:PS:FT:GQ:HQ:EHQ:CGA_CEHQ:GL:CGA_CEGL:DP:AD:CGA_RDP   1/0:.:PASS:
10     96448380  .       C      G      .     .      NS=1;AN=2;AC=1;CGA_XR=d
GT:PS:FT:GQ:HQ:EHQ:CGA_CEHQ:GL:CGA_CEGL:DP:AD:CGA_RDP   1/0:.:PASS:247:24
10     96448986  .       TTTAA  .      .     .      NS=1;AN=0    GT:PS  ./.:.
10     96448994  .       A      C      .     .      NS=1;AN=2;AC=1;CGA_FI=1
GT:PS:FT:GQ:HQ:EHQ:CGA_CEHQ:GL:CGA_CEGL:DP:AD:CGA_RDP   1|0:96448994:VQLO
10     96449112  .       TTTCTATATTTTTATGAA .      .      .      NS=1;AN=0
10     96449153  .       GTGGTCCTAGAAGGGACTGCTGC  .      .      .      NS=1;
10     96449429  .       A      .      .     .      NS=1;AN=0    GT:PS  ./.:.
10     96450001  .       T      <CGA_CNVWIN> .     .      NS=1;CGA_WINEND=9
10     96450204  .       ATTTCTCGACA  .      .     .      NS=1;AN=0    GT:PS
10     96450401  .       A      .      .     .      NS=1;AN=0    GT:PS  ./.:.
10     96450501  .       ATCCTTGGTGA  .      .     .      NS=1;AN=0    GT:PS
10     96450590  .       AGTAGGC      .      .     .      NS=1;AN=0    GT:PS
10     96450619  .       GGTCCAACATATGAAAATCAATAAAAGTAATCCAGTATA      .
10     96450677  .       AAACCACATGATTATCTCAACAGATGCAGAAAAGGCC .      .
10     96450743  .       AAAAACTCAATAAATTAGGTATTGATAGGA .      .      .
```

Converting VCF to FHIR



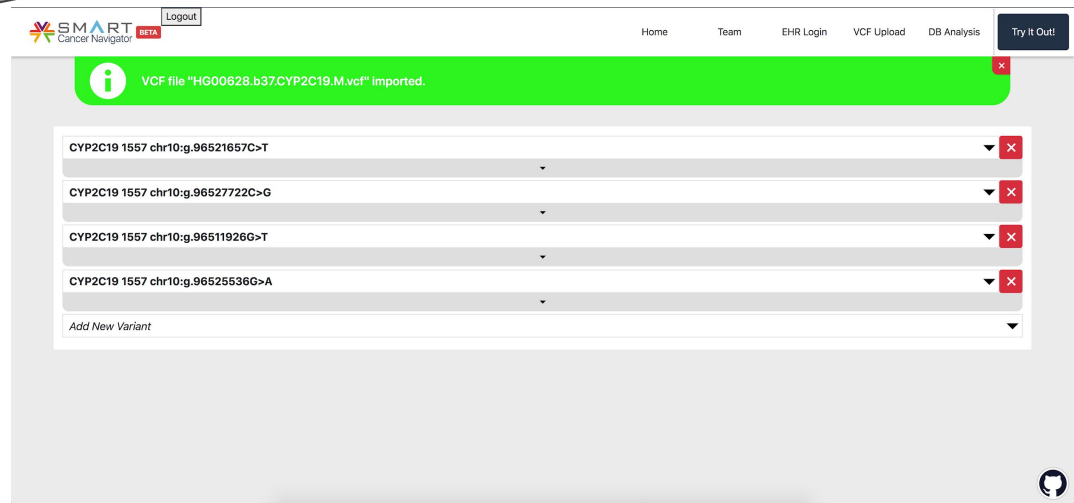

```
{
    "resourceType": "DiagnosticReport",
    "id": "dr-9194a2b6ee654",
    "meta": {
        "profile": [
            "http://hl7.org/fhir/uv/genomics-reporting/StructureDefinition/genomics-report"
        ]
    },
    "contained": [
        {
            "resourceType": "Observation",
            "id": "rs-0dadb95a69b54",
            "meta": {
                "profile": [
                    "http://hl7.org/fhir/uv/genomics-reporting/StructureDefinition/region-studied"
                ]
            },
            "status": "final",
            "category": [
                {
                    "coding": [
                        {
                            "system": "http://terminology.hl7.org/CodeSystem/observation-category",
                            "code": "laboratory"
                        }
                    ]
                }
            ],
            "code": {
                "coding": [
                    {
                        "system": "http://loinc.org",
                        "code": "53041-0",
                        "display": "DNA region of interest panel"
                    }
                ]
            },
```

```
{
  "resourceType": "DiagnosticReport",
  "id": "dr-9194a2b6ee654",
  "meta": {
    "profile": [
      "http://hl7.org/fhir/uv/genomics-reporting/StructureDefinition/genomics-report"
    ]
  },
  "contained": [
    {
      "resourceType": "Observation",
      "id": "rs-0dadb95a69b54",
      "meta": {
        "profile": [
          "http://hl7.org/fhir/uv/genomics-reporting/StructureDefinition/region-studied"
        ]
      },
      "status": "final",
      "category": [
        {
          "coding": [
            {
              "system": "http://terminology.hl7.org/CodeSystem/observation-category",
              "code": "laboratory"
            }
          ]
        }
      ],
      "code": {
        "coding": [
          {
            "system": "http://loinc.org",
            "code": "53041-0",
            "display": "DNA region of interest panel"
          }
        ]
      },
```

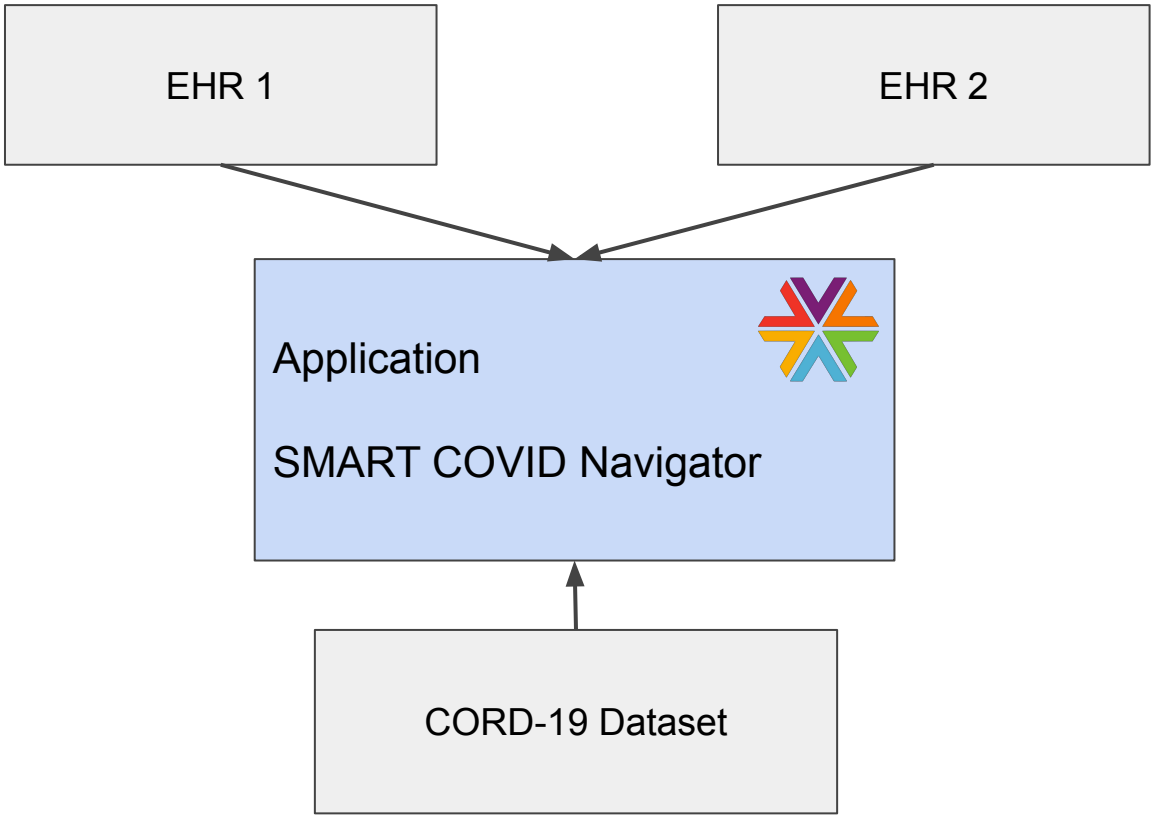

MyVariant.info

MyGene.info

Querying Gene-Variant Information



Phase 3: Extending Application to COVID-19



kaggle

CSV Headers		
Date	Severe Adjusted	Fatality Calculated
Study	Severe Calculated	Multivariate Adjustment
Study Link	Fatality	Study Type
Journal	Fatality lower bound	Sample Size
Severe	Fatality upper bound	Study Population
Severe lower bound	Fatality p-value	Added on
Severe upper bound	Fatality Significant	Critical only
Severe p-value	Fatality Adjusted	Discharged vs. death?
Severe Significant		

Risk Factors	
Age	Endocrine diseases
Asthma	Ethnicity: Hispanic vs. non-Hispanic
Autoimmune disorders	Heart Disease
COPD	Heart Failure
Cancer	Hypertension
Cardio- and cerebrovascular disease	Immune system disorders
Cerebrovascular disease	Male gender
Chronic digestive disorders	Neurological disorders
Chronic kidney disease	Overweight or obese
Chronic liver disease	Race: Asian vs. White
Chronic respiratory diseases	Race: Black vs. White
Dementia	Race: Other vs. White
Diabetes	Respiratory system diseases
Drinking	Smoking Status

SNOMED CT
The global language of healthcare

CMS
CENTERS FOR MEDICARE & MEDICAID SERVICES

ICD-10
ICD-9

VA: J96.11 Chronic respiratory failure with hypoxia
✓ VA: 53741008 Coronary Heart Disease
VA: G30.9 Alzheimer's disease, unspecified
VA: 22298006 Myocardial Infarction
VA: Z86.74 Personal history of sudden cardiac arrest
CMS: 70219 OTHER SBORHEIC KERATOSIS
CMS: 7099 SKIN DISORDER NOS
CMS: 7062 SEBACEOUS CYST
CMS: 7038 DISEASES OF NAIL NEC

Name: Gilberto712 Iglesias873 | **Zip Code:** 38004 | **Age:** 94 | **Condition:**

VA: 15777000 Prediabetes

Auto-Sync

Conditions

Age | Asthma | Autoimmune disorders | Cancer | Cardio- and cerebrovascular disease | Cerebrovascular disease

Chronic digestive disorders | Chronic kidney disease | Chronic liver disease | Chronic respiratory diseases | COPD

Dementia | Diabetes | Drinking | Endocrine diseases | Ethnicity_ Hispanic vs. non-Hispanic | Heart Disease

Heart Failure | Hypertension | Immune system disorders | Male gender | Neurological disorders

Overweight or obese | Race_ Asian vs. White | Race_ Black vs. White | Race_ Other vs. White

Respiratory system diseases | Smoking Status

VA: 157

Auto-Sync

Conditions

Age | Asthma | vascular disease | Chronic digestive | ases | COPD | Dementia | Dia | art Disease | Heart Failure | Overweight or ob | Respiratory syste

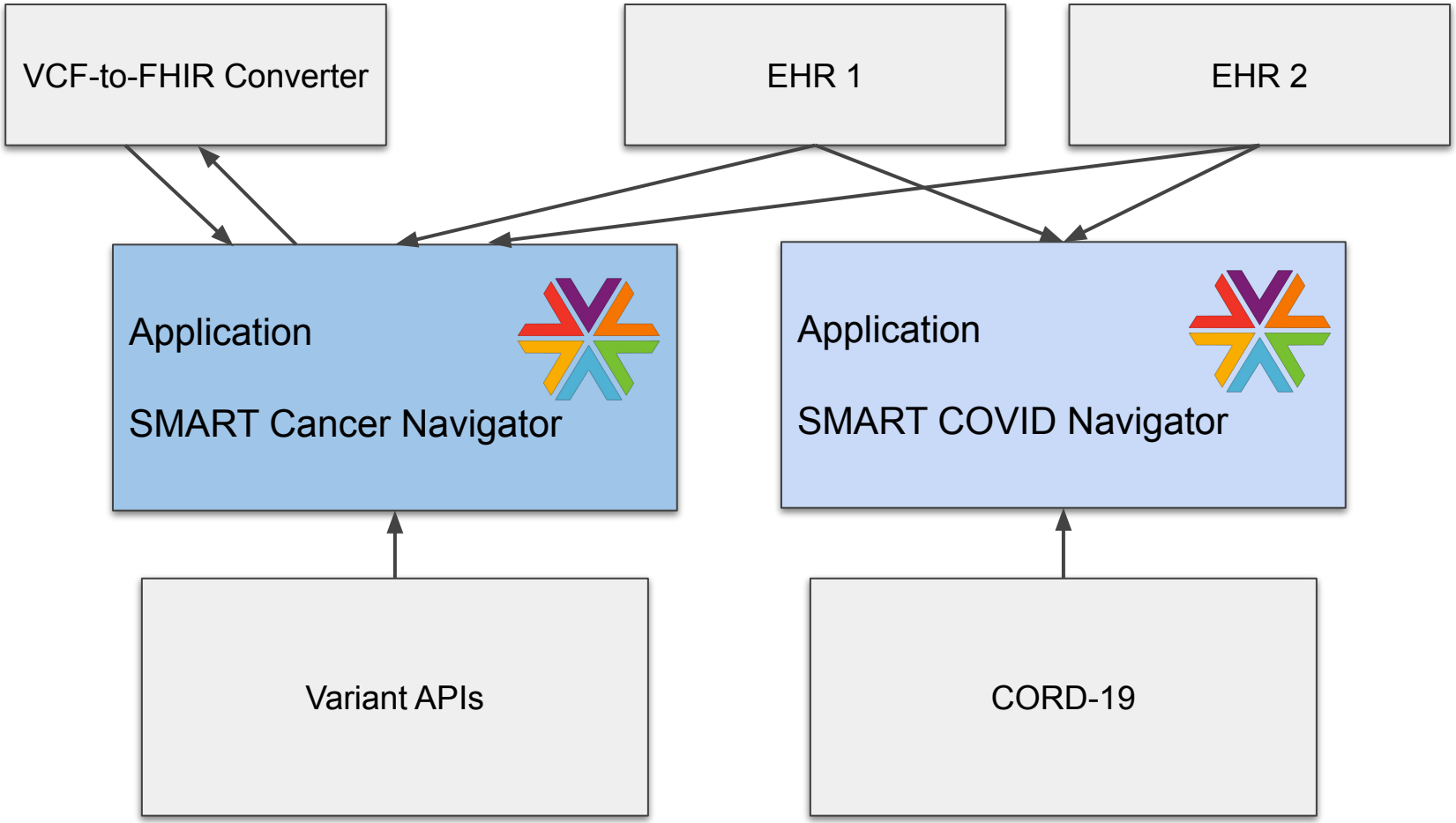
Age

Number of papers: 33

Proportion of significant findings for severity: 67%

Proportion of significant findings for fatality: 84%

Study Name	Date	Severity Significance	Fatality Significance
Extent of prior lung irradiation and mortality in COVID-19 patients with a cancer history	2020-05-20		Not Significant
Clinical Characteristics and Outcomes of Patients With Diabetes and COVID-19 in Association With Glucose-Lowering Medication	2020-05-14	Significant	Significant
Clinical Characteristics and Outcomes of Patients With Diabetes and COVID-19 in Association With Glucose-Lowering Medication	2020-05-14	Significant	Significant
Correlation of coagulation parameters with clinical outcomes in Coronavirus-19 affected minorities in United States: Observational cohort	2020-05-06		Significant


VCF-to-FHIR Converter
EHR 1
EHR 2
Application
SMART Cancer Navigator
Application
SMART COVID Navigator
Variant APIs
CORD-19

Future Work

- Add Reboot Rx COVID-Cancer Dataset
- Add AI/machine learning to cancer prediction using genomic data
- Add AI/machine learning to improve COVID-19 severity and mortality predictions

Acknowledgements

- MIT PRIMES
- Dr. Gil Alterovitz, Ning Xie, and Ling Teng
- Past PRIMES students who helped me understand prior work
- My family

Thanks!