GEL APIs

ACGS Bioinformatics Group Meeting

Aled Jones

12th June 2017

Application Programming Interface (API)

A way of accessing and interacting with an application

Interact using the url eg

https://bioinfo.extge.co.uk/crowdsourcing/WebServices/get_panel/56fa8eb88f62030f36e3026b/

PanelID

Data can be interrogated further

https://bioinfo.extge.co.uk/crowdsourcing/WebServices/get_panel/56fa8eb88f62030f36e3026b/?LevelOfConfidence=HighEvidence

PanelID Green genes

Data can be presented as JSON

https://bioinfo.extge.co.uk/crowdsourcing/WebServices/get_panel/56fa8eb88f62030f36e3026b/?LevelOfConfidence=HighEvidence/?format=json

PanelID Green genes JSON

GEL APIs include:

- PanelApp (Gene panels)
- CIP API (Referrals and test progress)
- LabKey (Patient demographics)

Python requests module

Built in package
Can read (GET) or submit (PUSH) data
Can submit authorisation token and proxy details

```
response = requests.get(url , headers={"Authorization":"JWT "+token})
```

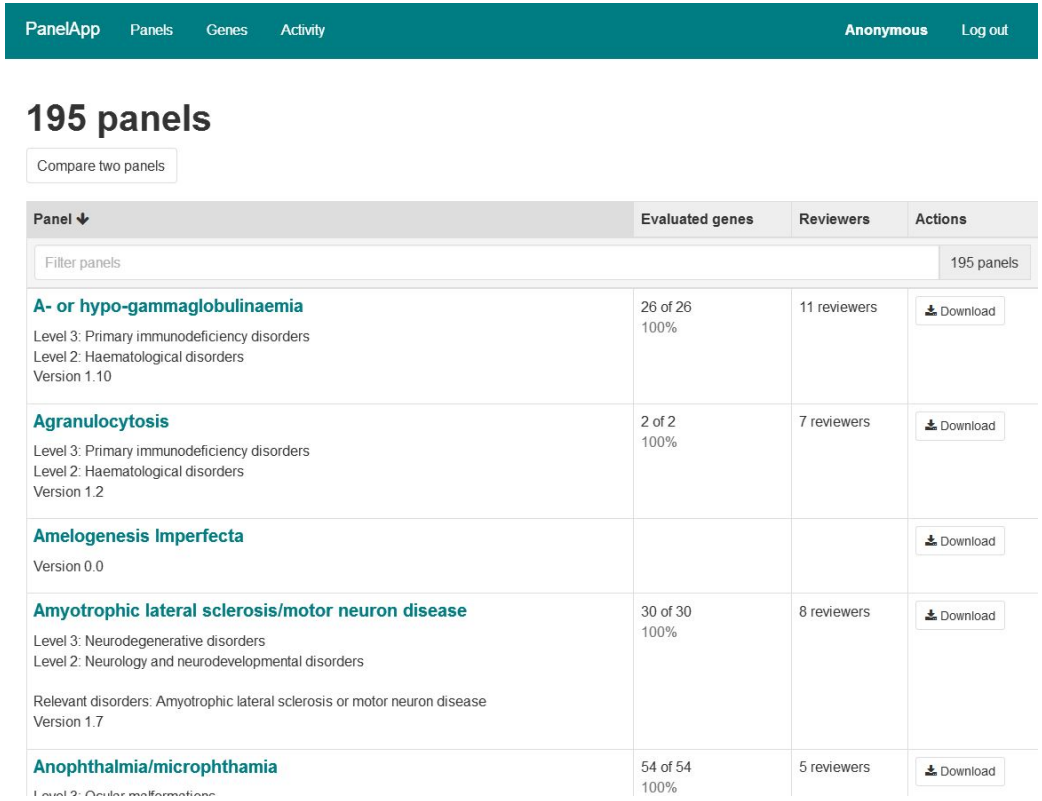
Returns an object you can parse

PanelApp

Crowd sourced curation of gene panels

Who is using the PanelApp panels/API???

PanelApp - 195 Panels



PanelApp - Each Panel has...

A ‘stable’ identifier

55d30b0322c1fc2ff2a5bf7b

A human readable name

Amyotrophic lateral sclerosis/motor neuron disease

A version number

v1.7 (only panels with a version number >= 1.0 are used to tier variants)

Green, Amber and Red Genes

Green = ‘High confidence’

Amber = ‘Intermediate confidence’

Red = ‘Low confidence’

PanelApp - How we use it

Incorporated panels into our LIMS to allow these to be selected as virtual panels in WES ordering system

1. Download all Green and Amber genes from all panels
   https://github.com/NHS-NGS/PanelApp_API
2. Import into our LIMS system
   Use stable panel ID (panel name can change and contains spaces,slashes etc)
   If panel has been updated…
      Add the green and amber gene panels
      Linking each gene to our internal HGNC snapshot.

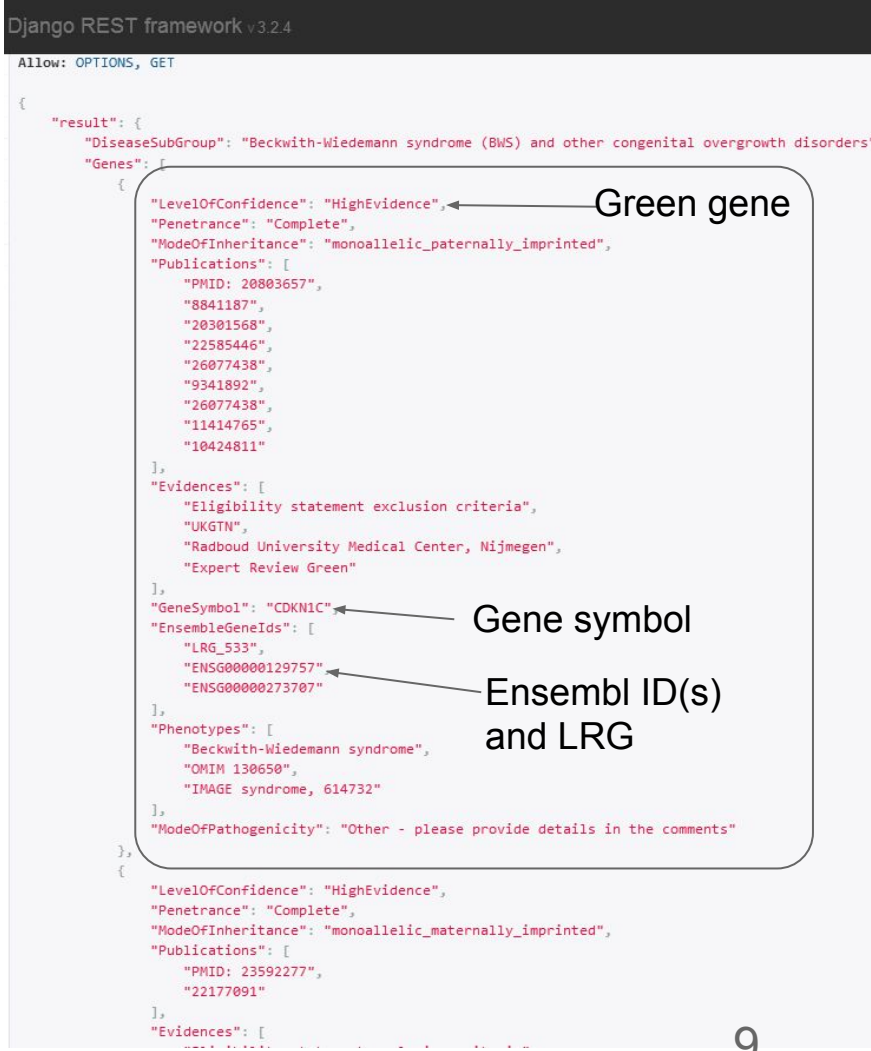
PanelApp - Problems

Gene Symbol - not stable
GEL HGNC snapshot quite old.

Our HGNC snapshot has 1 ensembl id per gene (?use http://mygene.info/)

List of ensembl ids has included ensembl ids for other genes

API != PanelApp website
additional identifiers (OMIM) are available via web



PanelApp - Our approach

We use the ensembl ids to join to our HGNC snapshot when importing genes.

We use the list of gene symbols to ensure we haven't missed any genes or imported incorrect genes.

Has required a (painful) manual curation of a 'translation' table between GEL gene symbols/ensembl ids and those in our HGNC snapshot. (Happy to share this)

CIP API

Clinical Interpretation Portal

Is anyone using the CIP-API???

CIP API - What is available

Each proband sent to GEL is given a participantID

Once analysed each proband has a JSON entry including:

- Links to download any files created, eg coverage reports

CIP API - What is available

Each proband sent to GEL is given a participantID

Once analysed each proband has a JSON entry including:

- Links to download any files created, eg coverage reports
- A status


```
"results": [
    {
        "files": [
            {
                "url": "http://cipapi.genomicsengland.nhs.uk/api/Interpretation/Download_file/1487252018.81/",
                "file_name": "LP2000857-DNA_C06.bw_55940f0e22c1fc4f7d26e965_1.1.json.gz",
                "file_type": "csv/"
            },
            {
                "url": "http://cipapi.genomicsengland.nhs.uk/api/Interpretation/Download_file/1487252017.77/",
                "file_name": "LP2000855-DNA_B07.bw_55940f0e22c1fc4f7d26e965_1.1.json.gz",
                "file_type": "csv/"
            }
        ],
        "number_of_samples": 2,
        "sites": [
            "GSTT"
        ],
        "last_status": "report_sent",
        "interpreted_genomes": [
            {
                "clinical_reports": [],
                "create_at": "2017-01-16T14:12:23.324345Z",
                "cip_version": 2
            },
            {
                "clinical_reports": [
                    {
                        "url": "http://cipapi.genomicsengland.nhs.uk/api/ClinicalReport/192/1/3/1/"
                    }
                ],
                "create_at": "2016-08-15T20:49:23.219118Z",
                "cip_version": 3
            },
            {
                "clinical_reports": [],
                "create_at": "2017-01-16T14:12:23.134326Z",
                "cip_version": 1
            }
        ],
        "proband": "50004182",
        "interpretation_request_id": "192-1",
        "sample_type": "raredisease",
        "last_update": "2017-05-09 12:27:26",
        "family_id": "50004182",
        "cip": "omicia"
    },
```

CIP API - What is available

Each proband sent to GEL is given a participantID

Once analysed each proband has a JSON entry including:

- Links to download any files created, eg coverage reports
- A status
- Clinical reports (one per cip version)


```
"results": [
    {
        "files": [
            {
                "url": "http://cipapi.genomicsengland.nhs.uk/api/Interpretation/Download_file/1487252018.81/",
                "file_name": "LP2000857-DNA_C06.bw_55940f0e22c1fc4f7d26e965_1.1.json.gz",
                "file_type": "csv/"
            },
            {
                "url": "http://cipapi.genomicsengland.nhs.uk/api/Interpretation/Download_file/1487252017.77/",
                "file_name": "LP2000855-DNA_B07.bw_55940f0e22c1fc4f7d26e965_1.1.json.gz",
                "file_type": "csv/"
            }
        ],
        "number_of_samples": 2,
        "sites": [
            "GSTT"
        ],
        "last_status": "report_sent",
        "interpreted_genomes": [
            {
                "clinical_reports": [],
                "create_at": "2017-01-16T14:12:23.324345Z",
                "cip_version": 2
            },
            {
                "clinical_reports": [
                    {
                        "url": "http://cipapi.genomicsengland.nhs.uk/api/ClinicalReport/192/1/3/1/"
                    }
                ],
                "create_at": "2016-08-15T20:49:23.219118Z",
                "cip_version": 3
            },
            {
                "clinical_reports": [],
                "create_at": "2017-01-16T14:12:23.134326Z",
                "cip_version": 1
            }
        ],
        "proband": "50004182",
        "interpretation_request_id": "192-1",
        "sample_type": "raredisease",
        "last_update": "2017-05-09 12:27:26",
        "family_id": "50004182",
        "cip": "omicia"
    },
```

Proband Participant ID: 50004237 Generated on: 02 May 2017 16:43

Whole Genome Analysis
Rare Disease Primary Findings

Genomics england

Genomics England, Queen Mary University of London,
Dawson Hall, Charterhouse Square, EC1M 6BQ.
Email: GENOMICS.pilot-results@nhs.net

NHS Genomic Medicine Centre: **GSTT**
Interpretation Request ID: **OPA-431-2**

Participant Information

Family ID	Gel Participant ID	Sample	Year of Birth	Gender	Relationship to Proband	Disease Status
[redacted]						

Human Phenotype Ontology (HPO) terms for the Proband

Positive: Cardiac arrest, Adrenal pheochromocytoma, Gangrene

Gene Panels

Gene Panel Name	Gene Panel Version	Gene Panel Size
Multiple endocrine tumours	1.3	16
Neuro-endocrine Tumours- PCC and PGL	1.0	20

Report context

Genomics England Primary Findings describes candidate variant(s) selected by the recruiting Genomics Medicine Centre from the perspective of the participant. It may include variants in the virtual gene panel(s) classified as tier 1 or 2 and/ or additional prioritised variants of relevance to the patient's phenotype. Primary findings do not include additional looked-for findings unrelated to the patient's phenotype. The tiering has been carried out based on the phenotype and pedigree data as given in this report. It is the responsibility of the reporting accredited laboratory to check that this information is correct before issuing a patient report.

Tiered variants are rare variants that segregate with disease. Tier 1: Includes high impact variants (stop-gained, stop-lost, start lost, splice donor/acceptor, frameshift, transcript ablation, or de novo in a monoallelic gene) within a curated list of genes with sufficient evidence associating them with the patient's phenotype(s). Tier 2: Includes moderate impact variants (missense, splice region variant, in-frame insertion/deletion, transcript amplification, incomplete terminator codon) within a curated list of genes with sufficient evidence associating them with the patient's phenotype(s). Tier 3: Includes high and moderate impact variants outside of the curated list of genes associated with the patient's phenotype(s). The large majority of Tier 3 variants will NOT be pathogenic.

Unless otherwise stated, this report describes unvalidated variant predictions, not diagnostic results. This report only contains information about single nucleotide variants and small insertions/deletions; copy number variants and structural variants have NOT yet been analysed. It is possible that disease-causing variant(s) are located outside of this list of prioritised variants, for example because they fall outside the gene panels applied, because they were located in regions of low coverage or because the variant is of a type that could not be called. Further diagnostic or research analysis may lead to an updated report being issued in the future.

Variants

⚠ No variants found

Genome Intepretation

REPORT ANNEX

Coverage Metrics (Click to collapse/expand)

panel: Multiple endocrine tumours

Format: [**gene symbol:** *Participant Id*(**proportion of sample covered greater than 15x [calculated using union of transcripts, coding regions +/- 15bp]**) ...]
AIP: *50004237*:**(0.96)**
CDC73: *50004237*:**(1.00)**
CDKN1B: *50004237*:**(0.99)**
MAX: *50004237*:**(0.99)**
MEN1: *50004237*:**(0.97)**
PRKAR1A: *50004237*:**(1.00)**
RET: *50004237*:**(0.99)**
SDHAF2: *50004237*:**(1.00)**
SDHB: *50004237*:**(1.00)**
SDHC: *50004237*:**(0.92)**
SDHD: *50004237*:**(0.97)**
SUMMARY: *50004237*:**(0.99)**
TMEM127: *50004237*:**(0.99)**
TP53: *50004237*:**(0.99)**
VHL: *50004237*:**(0.99)**

panel: Neuro-endocrine Tumours- PCC and PGL

Format: [**gene symbol:** *Participant Id*(**proportion of sample covered greater than 15x [calculated using union of transcripts, coding regions +/- 15bp]**) ...]
FH: *50004237*:**(0.99)**
MAX: *50004237*:**(0.99)**
MEN1: *50004237*:**(0.97)**
NF1: *50004237*:**(0.99)**
PRKAR1A: *50004237*:**(1.00)**
RET: *50004237*:**(0.99)**
SDHA: *50004237*:**(0.97)**
SDHAF2: *50004237*:**(1.00)**
SDHB: *50004237*:**(1.00)**
SDHC: *50004237*:**(0.92)**
SDHD: *50004237*:**(0.97)**
SUMMARY: *50004237*:**(0.99)**
TMEM127: *50004237*:**(0.99)**
VHL: *50004237*:**(0.99)**

Referenced Databases

Name	Version
ClinVar	2016-07-11
GWAS	2014-05-21
1000 Genomes Project	Phase 3 Version 5 (2013-05-02)
ExAC	v0.3
GRCh	v37
HGMD	Professional 2016.2
OMIM	2016-07-11
GeL Frequency	October 2016
CADD	v1.0
Ensembl	v83
dbSNP	v147

Software Versions

Name	Version
PhyloP	phyloP46way (2009-12-01)
phevor_version	2.1.1
MutationTaster	dbNSFP v2.9
VVP	v1.1
PolyPhen-2	v2.2.2 (dbNSFP v2.9)
SIFT	dbNSFP v2.9
Omicia Score	v2.0
vaast_version	3.0.4.2
pipeline_version	6.0.4

CIP API - What is available

Each proband sent to GEL is given a participantID

Once analysed each proband has a JSON entry including:

- Links to download any files created, eg coverage reports
- A status
- Clinical reports (one per cip version)
- Participant and other IDs


```
"results": [
    {
        "files": [
            {
                "url": "http://cipapi.genomicsengland.nhs.uk/api/Interpretation/Download_file/1487252018.81/",
                "file_name": "LP2000857-DNA_C06.bw_55940f0e22c1fc4f7d26e965_1.1.json.gz",
                "file_type": "csv/"
            },
            {
                "url": "http://cipapi.genomicsengland.nhs.uk/api/Interpretation/Download_file/1487252017.77/",
                "file_name": "LP2000855-DNA_B07.bw_55940f0e22c1fc4f7d26e965_1.1.json.gz",
                "file_type": "csv/"
            }
        ],
        "number_of_samples": 2,
        "sites": [
            "GSTT"
        ],
        "last_status": "report_sent",
        "interpreted_genomes": [
            {
                "clinical_reports": [],
                "create_at": "2017-01-16T14:12:23.324345Z",
                "cip_version": 2
            },
            {
                "clinical_reports": [
                    {
                        "url": "http://cipapi.genomicsengland.nhs.uk/api/ClinicalReport/192/1/3/1/"
                    }
                ],
                "create_at": "2016-08-15T20:49:23.219118Z",
                "cip_version": 3
            },
            {
                "clinical_reports": [],
                "create_at": "2017-01-16T14:12:23.134326Z",
                "cip_version": 1
            }
        ],
        "proband": "50004182",
        "interpretation_request_id": "192-1",
        "sample_type": "raredisease",
        "last_update": "2017-05-09 12:27:26",
        "family_id": "50004182",
        "cip": "omicia"
    },
```

CIP API - Generating clinical reports

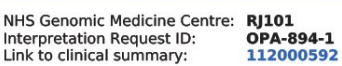
GitHub repo
https://github.com/NHS-NGS/GEL_reports

- Python script which reads API
- Downloads the GEL html report (most recent report from highest CIP version)
- Modifies html
- Produces PDF

CIP API - Generating clinical reports

- Replace/remove GeL Logo and address to make it obvious that this is not a GeL report

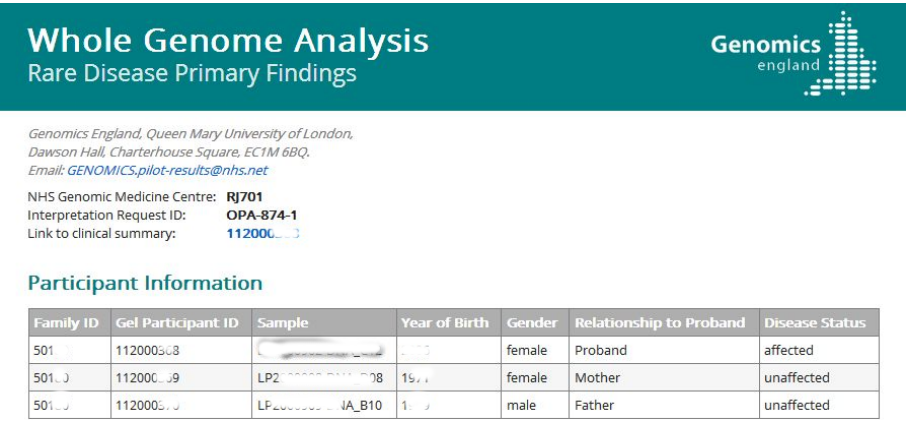
Clinician name
Clinician address
100,000 genomes project result



viapath

CIP API - Generating clinical reports

- Proband information added from LIMS



Whole Genome Analysis
Rare Disease Primary Findings

Genomics england

Genomics England, Queen Mary University of London,
Dawson Hall, Charterhouse Square, EC1M 6BQ.
Email: GENOMICS.pilot-results@nhs.net

NHS Genomic Medicine Centre: **RJ701**
Interpretation Request ID: **OPA-874-1**
Link to clinical summary: **112000**

Participant Information

Family ID	Gel Participant ID	Sample	Year of Birth	Gender	Relationship to Proband	Disease Status
501	1120003 8			female	Proband	affected
501	112000 9	LP2 08	19	female	Mother	unaffected
501	112000	LP A_B10	1	male	Father	unaffected

```
Clinician name
Clinician address
      100,000 genomes project result
```

viapath

NHS Genomic Medicine Centre: **RJ101**
Interpretation Request ID: **OPA-894-1**
Link to clinical summary: **112000592**

Proband Information

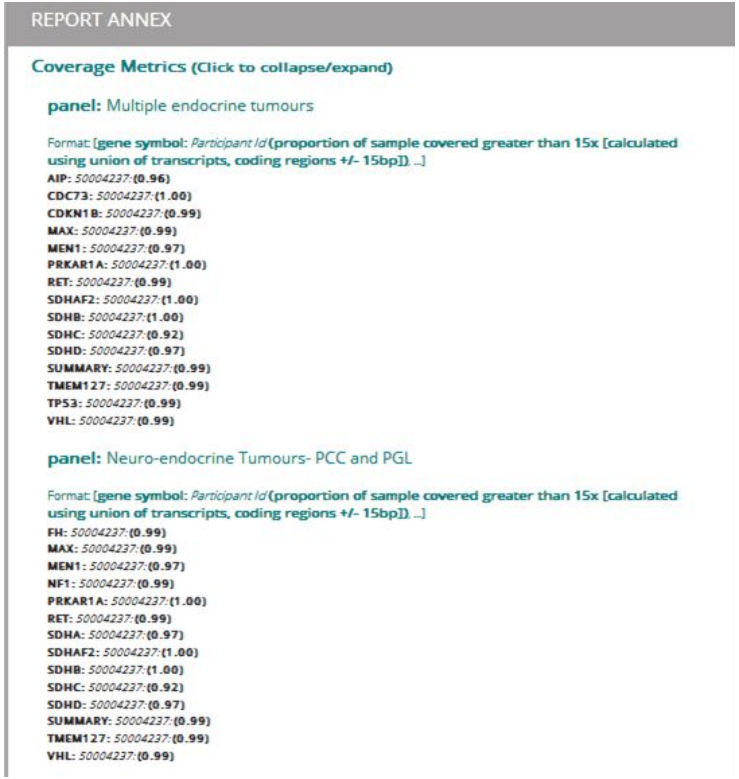
First Name	Last Name	Date Of Birth	Gender	PRU	NHS Number
{{firstname}}	{{lastname}}	{{dob}}	{{gender}}	{{PRU}}	{{NHS}}

Participant Information

Family ID	Gel Participant ID	Sample	Year of Birth	Gender	Relationship to Proband	Disease Status
FM5 3	5 2	LP2 DNA 07	2011	female	Proband	affected
FM500 3	5 5	LP 7- DNA	1981	female	Mother	unaffected

CIP API - Generating clinical reports

- Coverage report expanded



REPORT ANNEX

Coverage Metrics (Click to collapse/expand)

panel: Multiple endocrine tumours

Format: [**gene symbol:** *Participant Id* (**proportion of sample covered greater than 15x [calculated using union of transcripts, coding regions +/- 15bp]**). ...]

AIP: *50004237*:(**0.96**)
CDC73: *50004237*:(**1.00**)
CDKN1B: *50004237*:(**0.99**)
MAX: *50004237*:(**0.99**)
MEN1: *50004237*:(**0.97**)
PRKAR1A: *50004237*:(**1.00**)
RET: *50004237*:(**0.99**)
SDHAF2: *50004237*:(**1.00**)
SDHB: *50004237*:(**1.00**)
SDHC: *50004237*:(**0.92**)
SDHD: *50004237*:(**0.97**)
SUMMARY: *50004237*:(**0.99**)
TMEM127: *50004237*:(**0.99**)
TP53: *50004237*:(**0.99**)
VHL: *50004237*:(**0.99**)

panel: Neuro-endocrine Tumours- PCC and PGL

Format: [**gene symbol:** *Participant Id* (**proportion of sample covered greater than 15x [calculated using union of transcripts, coding regions +/- 15bp]**). ...]

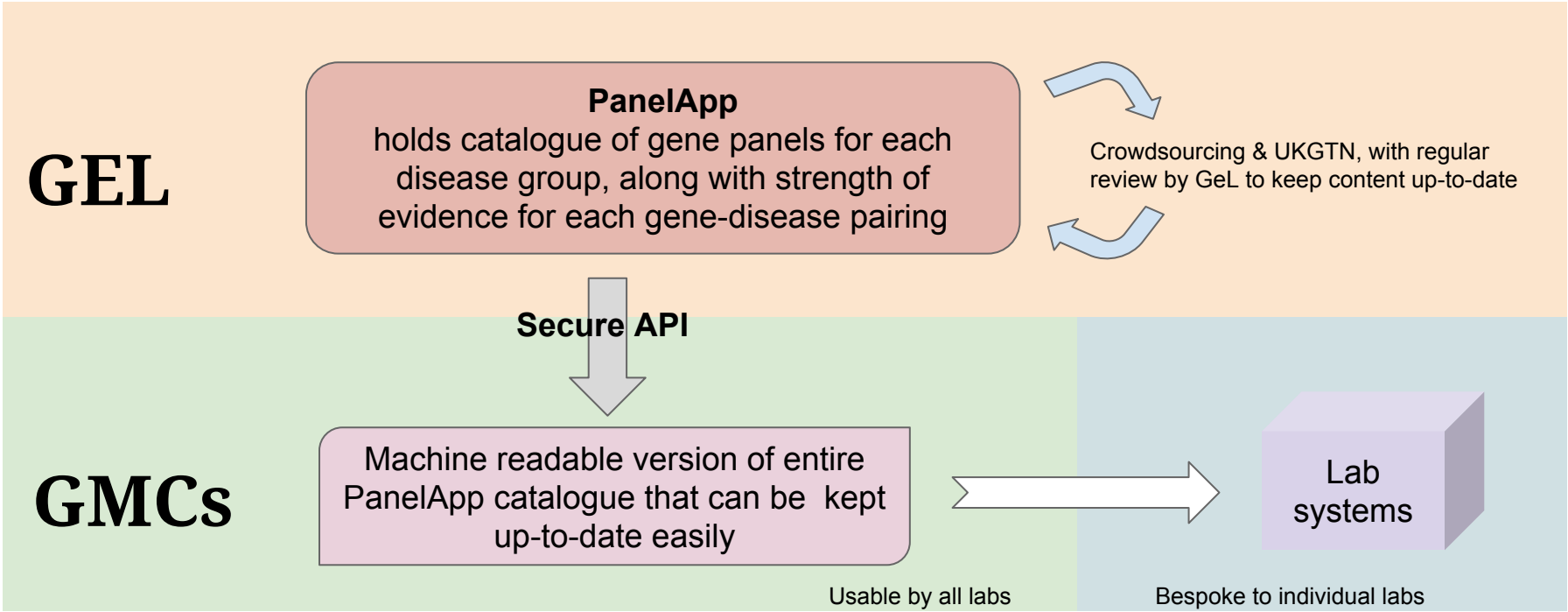
FH: *50004237*:(**0.99**)
MAX: *50004237*:(**0.99**)
MEN1: *50004237*:(**0.97**)
NF1: *50004237*:(**0.99**)
PRKAR1A: *50004237*:(**1.00**)
RET: *50004237*:(**0.99**)
SDHA: *50004237*:(**0.97**)
SDHAF2: *50004237*:(**1.00**)
SDHB: *50004237*:(**1.00**)
SDHC: *50004237*:(**0.92**)
SDHD: *50004237*:(**0.97**)
SUMMARY: *50004237*:(**0.99**)
TMEM127: *50004237*:(**0.99**)
VHL: *50004237*:(**0.99**)

CIP API - Generating clinical reports

Code written in a way that will (hopefully) be easily implemented by any lab.

- Config file
- Requirements and notes in ReadMe
- Patient info table can be modified as required.

Local implementation of PanelApp for selection of exome virtual panel tests

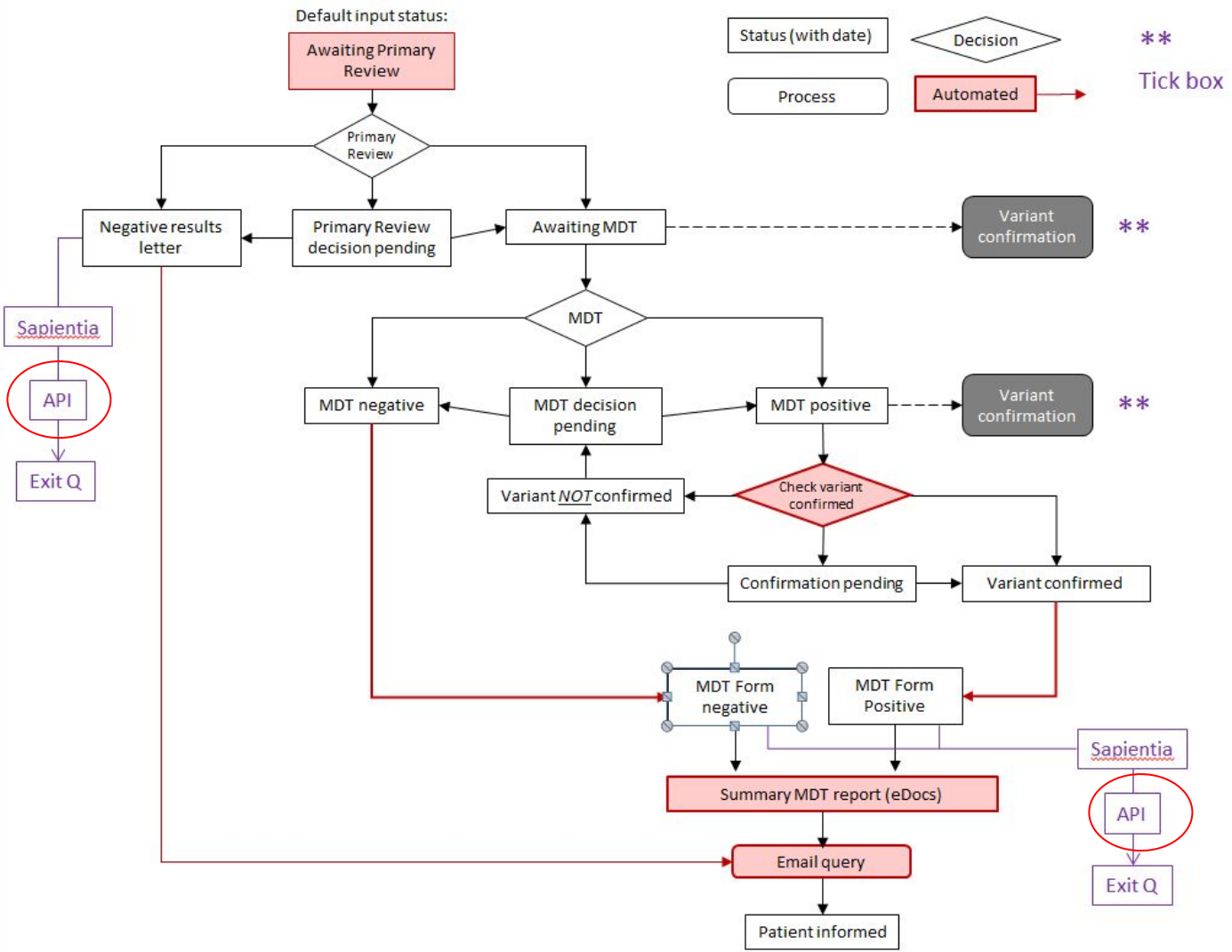


https://github.com/NHS-NGS/PanelApp_API

How else can we use the APIs???

- Track sample status/progress
- Labkey

Detail of one system



Indel benchmarking tool

Coming very soon

About

This tool is provided for clinical labs to assess the indel calling performance of their NGS pipelines. It uses the precisionFDA workflow (vcfeval + hap.py) to compare a sample VCF to a truth set. BAM files containing known indels are provided for you to run through your pipeline (this data has been simulated to represent indels that might be sequenced as part of an exome workflow, with read depth in the region of 20x). The resulting VCF can then be uploaded, and performance metrics will be reported back.

Instructions

1. Download BAM file (aligned with bwa mem) from **here**
2. Process BAM through pipeline
3. Submit the gzipped VCF file produced by your pipeline, along with your email address, using the form to the right
4. Await an email containing a link to your results

Useful Info

- PrecisionFDA guide to processing and results
- Workflow illustration

Contact

If you any questions or comments, please post in **Slack**

Submit

Email Address

Results will be sent to this address

Attach gzipped VCF (.vcf.gz)

Choose file No file chosen

Submit