ACGS

Association for Clinical Genomic Science

Bioinformatics group update

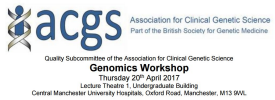
Joo Wook Ahn, Guy's & St Thomas'

26/06/2017 - ACGS Summer Meeting

Recent activity…

Bi-annual group meeting

Nov ‘16 - Bristol // June ‘17 - Leeds // Nov ‘17 - Oxford



ACGS Genomics Workshop

April ‘17 - Manchester

Bioinformatics “slack” channels

NHS bioinformaticians + GeL bioinformaticians

Code sharing

NHS-NGS github organisation

1.

Scope of roles

Is there a line between bioinformatician and clinical scientist?

Scientist expertise domains

As the scope of diagnostic test gets bigger, e.g. single gene tests -> WGS, it’s no longer feasible to have one person do everything.

1. Clinical genomics software engineer

Creates clinical-grade software to enable utilising NGS (and beyond)

Understands infrastructure

2. Clinical genomics data scientist

Determines how to best use data.

For example:

- reference genome data, e.g. appropriate use of gnomAD constraint scores
- genetic algorithms, e.g. is this estimated coefficient appropriate when prevalence is so low

3. Clinical genomics variation interpretation scientist

Interprets variation in context of phenotype & segregation etc, presents case to MDT and reports the results of testing

Clinical genomics roles

What's the most efficient way to have all expertise domains covered?

One person with many (all) skills?

Many people, each specialised in a few skills?

2.

Benchmarking

A standardised method to calculate sensitivity

Validation of NGS assays

- Sequence GIAB reference material
- Process data & call variants
- Assess sensitivity etc

Validation of NGS assays

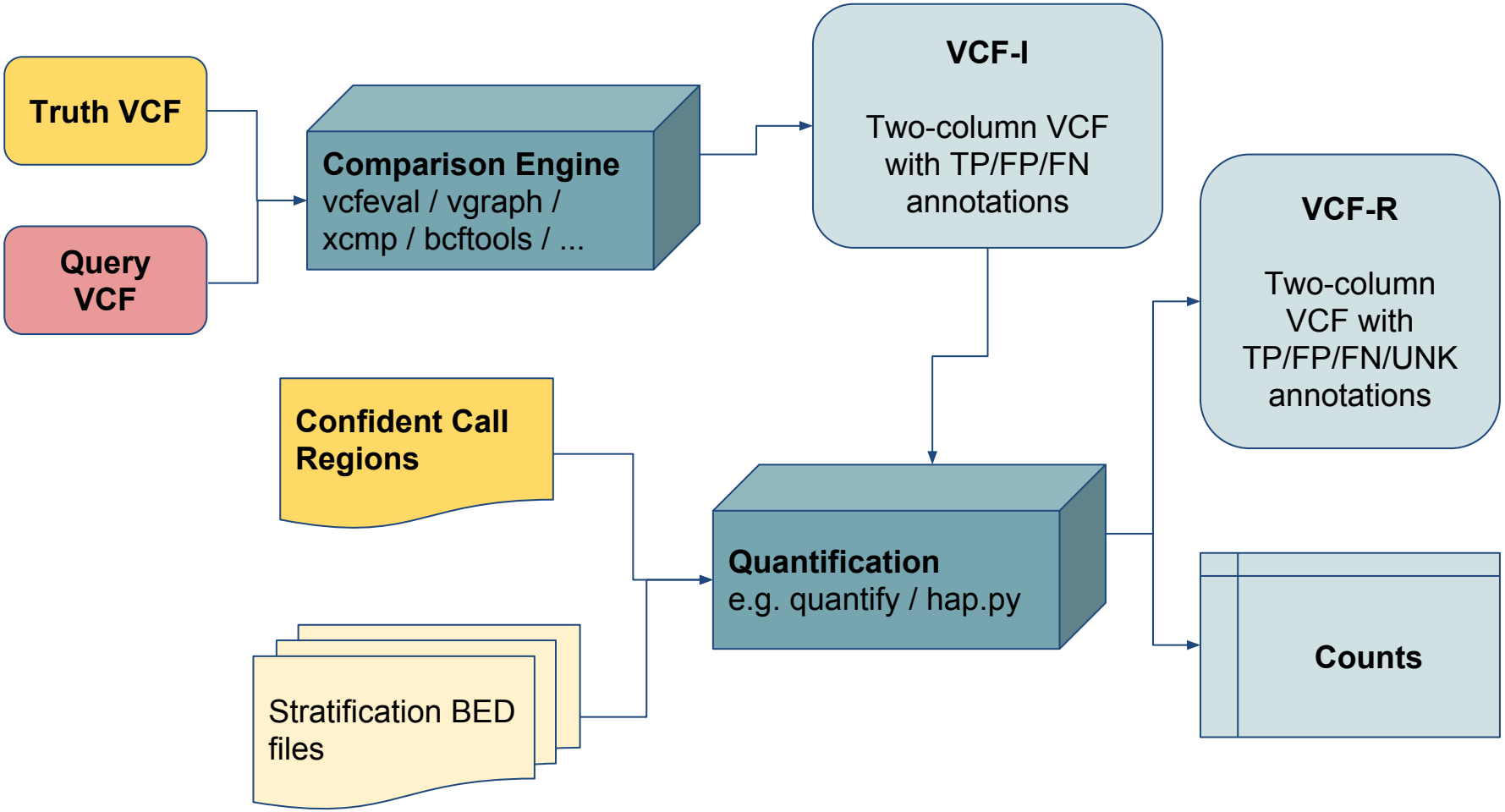
- Sequence GIAB reference material
- Process data & call variants
- Assess sensitivity etc
  - Not trivial
  - Different software can give different results

2016 benchmarking exercise

Variant caller	Number of indels missed (n=139)
GATK-lite (unified genotyper)	9
Samtools	7
NextGene	3 (+ 24 mis-annotated)
GATK v3	2
GATK v3	1
Platypus	0
GATK v3	0

Global Alliance for Genomics & Health

Benchmarking Tool Architecture



Source: Peter Krusche; https://github.com/ga4gh/benchmarking-tools

precisionFDA

A standardised tool

- Take PrecisionFDA implementation
- Adapt for NHS
- Make available to community

precisionFDA

A standardised tool

- Take PrecisionFDA implementation
- Adapt for NHS
- Make available to community

➢ **Inter-lab comparison possible**

http://stickie.be/nch

About

This tool is provided for clinical labs to assess the performance of their NGS workflows. It uses the precisionFDA workflow (vcfeval + hap.py) to compare a sample VCF to the NIST Genome in a Bottle NA12878 truth set. See links below for more information.

You can use this tool to assess NA12878 sequencing data you have generated by supplying a VCF file containing your variants and a BED file to restrict analysis to regions covered by your panel.

Alternatively if you just want to assess your bioinformatics pipeline, a pair of FASTQ files are provided. These were generated from Illumina NextSeq paired end whole exome sequencing on the NA12878 sample. You can run these through your pipeline and upload the resulting VCF (no BED file necessary).

You will receive an email containing the output from hap.py, which includes a summary file containing recall (sensitivity) and precision, as well as more detailed results. See the links below for more detail about interperting these results.

Useful Info

- PrecisionFDA guide to processing and results
- Workflow illustration
- Illumina hap.py
- RTG Tools vcfeval

Our Results

As a reference point we are provinding our results for the provided FASTQ files, which we ran through our pipeline using BWA MEM and GATK Best Practices. The summary is shown below, and the full result set can be downloaded here.

Type	Recall (Sensitivity)	Precision
SNPs	0.994763	0.996936
INDELs	0.91738	0.907096

Instructions

1. (OPTIONAL) Download paired end FASTQ files (~12GB): http://stickie.be/nch/mark/FASTQ/NA12878_WES.zip
2. Process above FASTQs or your own NA12878 data through your pipeline
3. Submit the gzipped VCF file produced by your pipeline, along with your email address and an optional BED file to restrict analysis regions, using the form below
4. Await an email containing a link to your results.
   WARNING: Email might be marked as spam, please whitelist gst-tr.mokaguys@nhs.net!!

Submit

Email Address

Results will be sent to this address

Attach gzipped VCF (.vcf.gz)

Choose File No file chosen

Attach analysis region BED (.bed) (OPTIONAL)

Choose File No file chosen

Submit

PLEASE NOTE: Upload may be slow if large VCF files are being submitted. Please be patient!

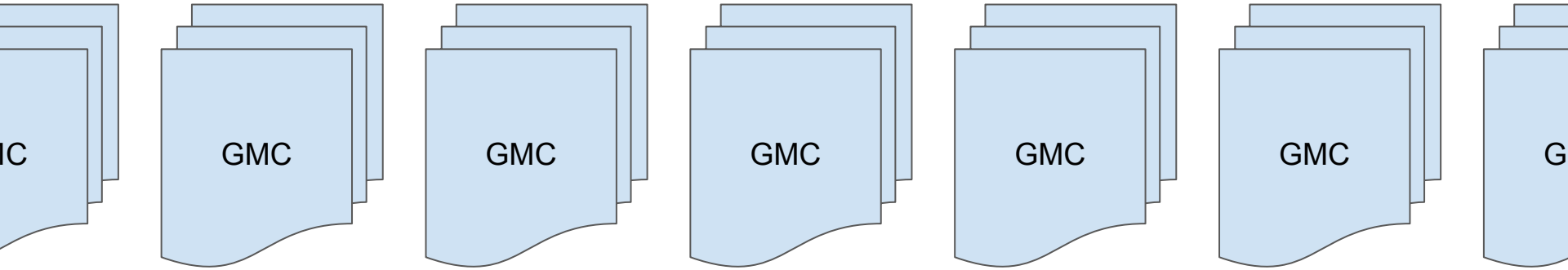
ACGS Association for Clinical Genomic Science

3.
GMC-GeL interfaces

100,000 genomes project,
genetics laboratories reconfiguration
…

Current landscape

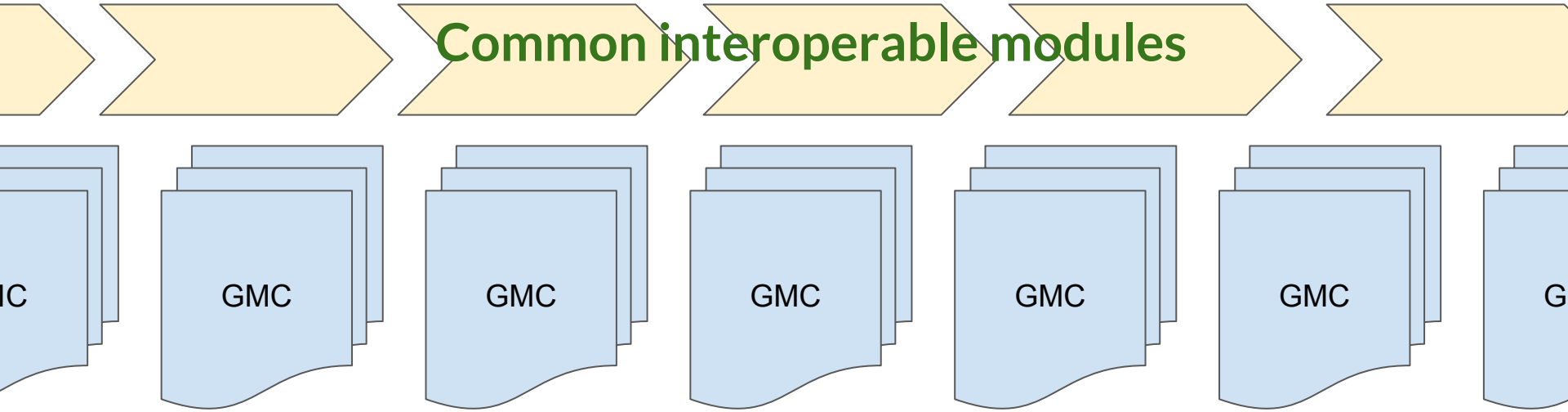
GeL systems & APIs
- “One” system
- APIs for programmatic access



Multiple disparate systems

Proposed landscape

GeL systems & APIs
- “One” system
- APIs for programmatic access

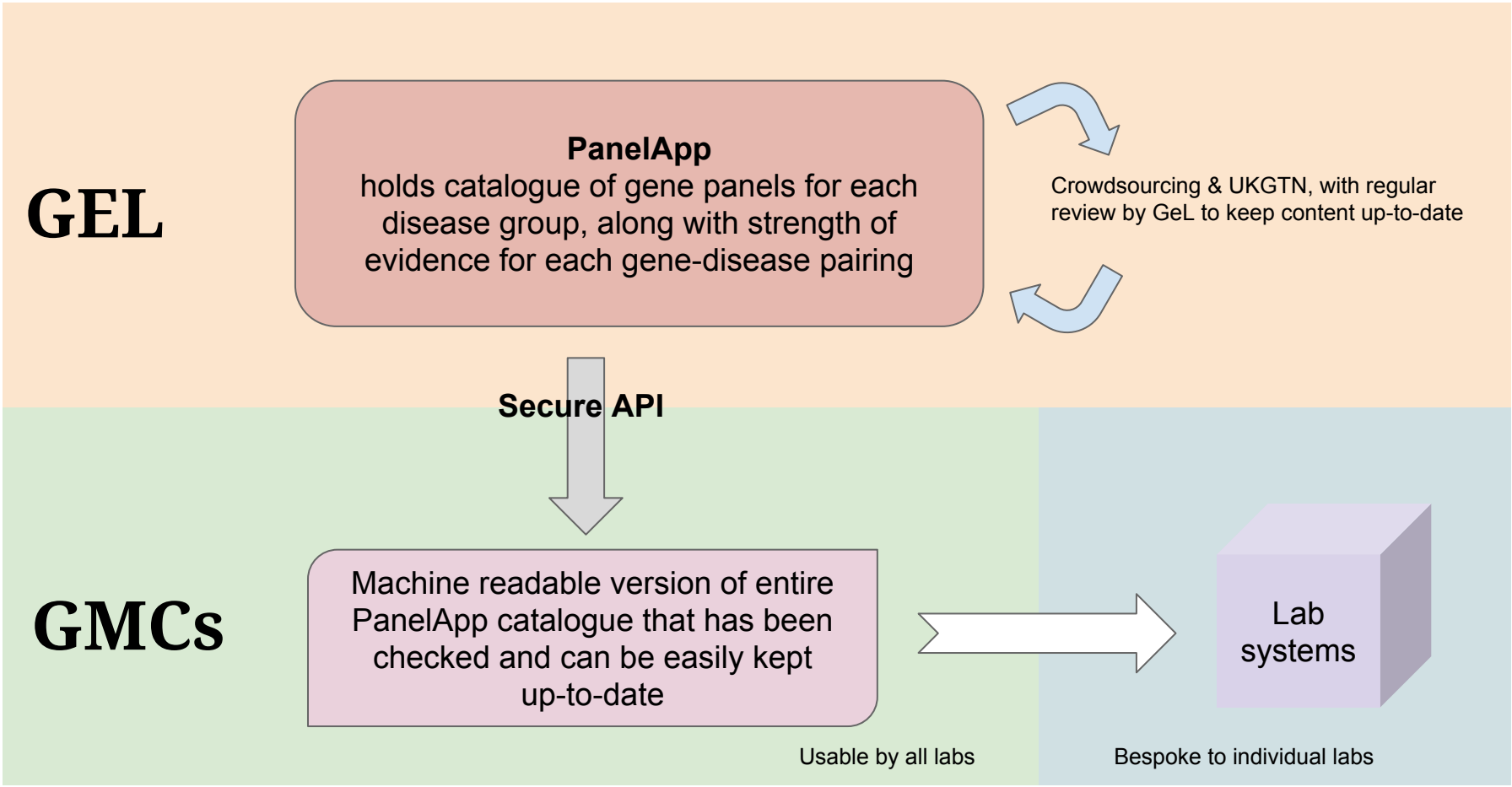


Multiple disparate systems

GMC :: GeL interfaces

- Develop as a collaborative effort between GMCs and GeL
- Develop with core functionality that's usable for all GMC labs
- Set ACGS policy to coordinate this effort to develop tools for working with GeL resources

Local implementation of PanelApp for selection of gene panel



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

Adding demographics to 100K report

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	112000308			female	Proband	affected
501	112000 09	LP2 08	19	female	Mother	unaffected
501	112000	LP NA_B10	1	male	Father	unaffected

Adding demographics to 100K report

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**

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	5	LP2 DNA_B07	2011	female	Proband	affected
FM500	5	LP DNA	1981	female	Mother	unaffected

Release due on https://github.com/NHS-NGS soon

ACGS

Association for Clinical Genomic Science

Pavlos Antoniou
Austin Diamond
Garan Jones
Kim Brugger
Kevin Ryan
Andrew Bond
Aled Jones
Jan Taylor
Chris Boustred
George Asimenos

GeL bioinformatics
NHS genetics labs

Thanks!