Genomics in the NHS Today

Best Practice

Matthew Parker, Ph. D
Lead Bioinformatician
Sheffield Diagnostic Genetics Service

Sheffield Children's NHS
NHS Foundation Trust

Introduction

Sheffield Children's NHS
NHS Foundation Trust

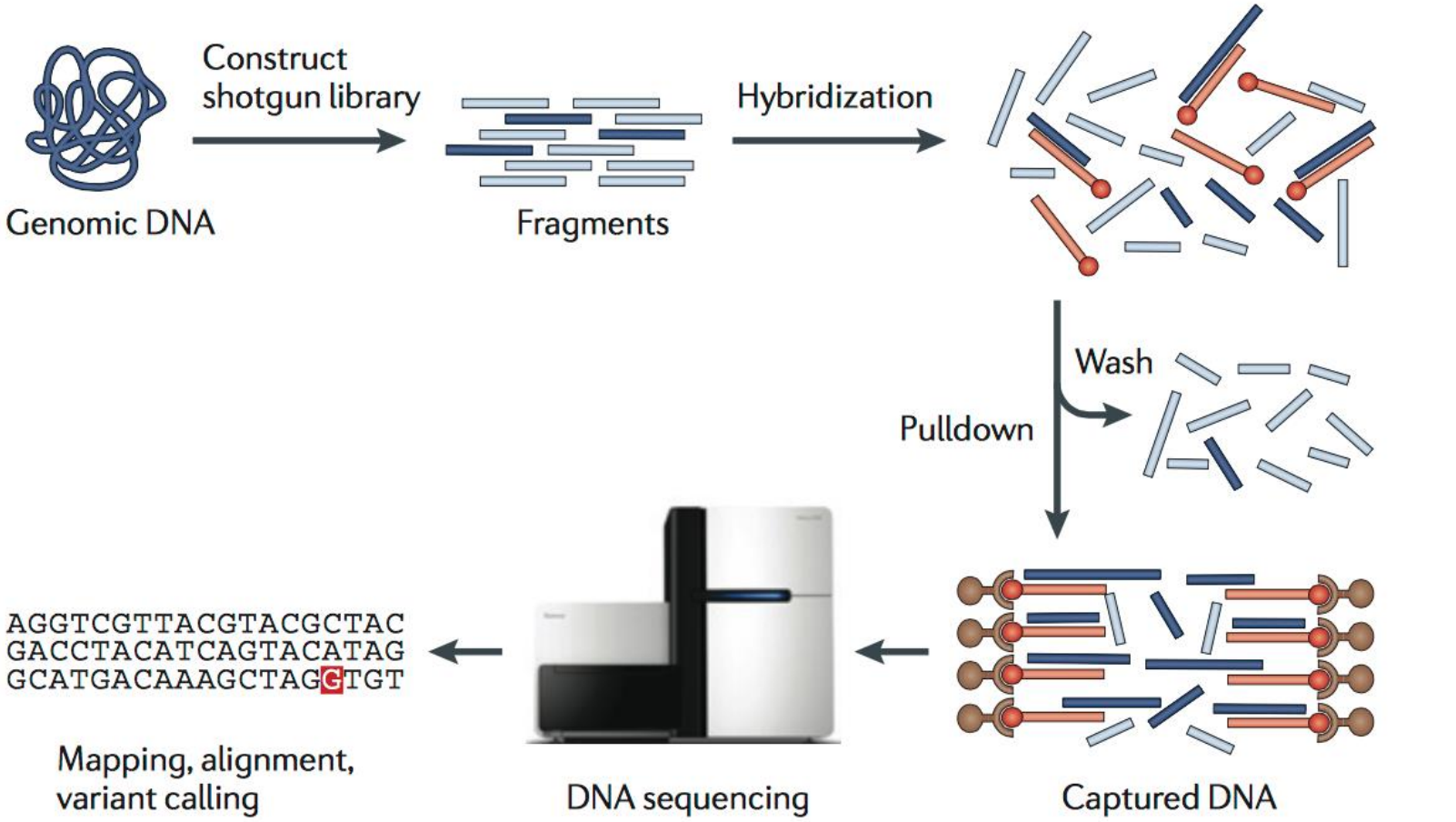
- NHS has around 40 full time bioinformaticians and trainees
- Most labs are sequencing around 3000 patients a year
- Around 10 labs – 30,000 patients a year
- Panels or exomes: Targeted panels introduced into NHS 2010 sequencing
- Disparate pipelines
- NHS model where labs compete for work
- Historically: Bioinformatics substitute for poor IT support & infra

SDGS Bioinformatics Team

Types of Sequencing in the NHS

Sheffield Children's NHS
NHS Foundation Trust

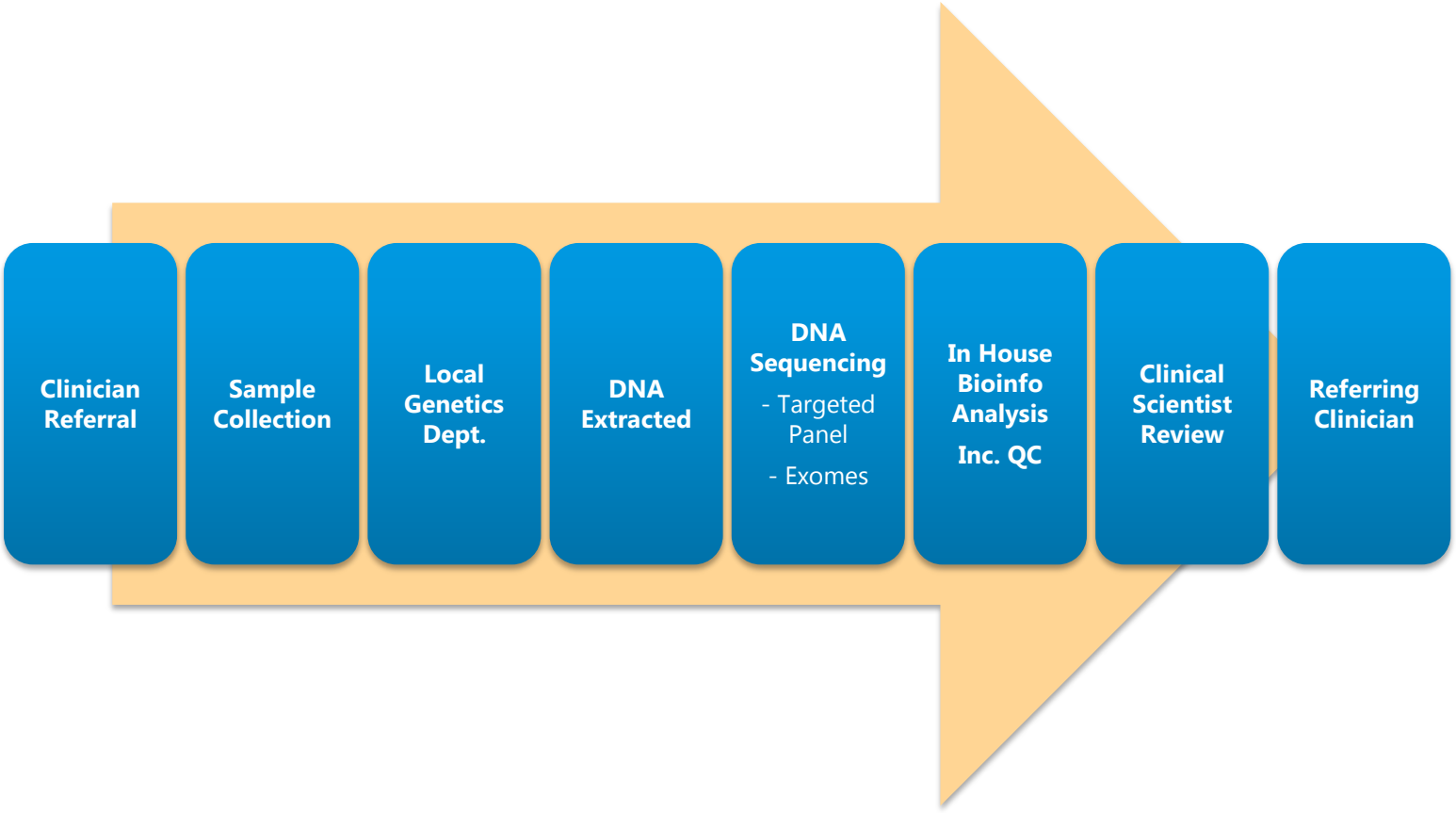
Targeted Panels or Exomes (GEL doing genomes)



NATURE REVIEWS | GENETICS

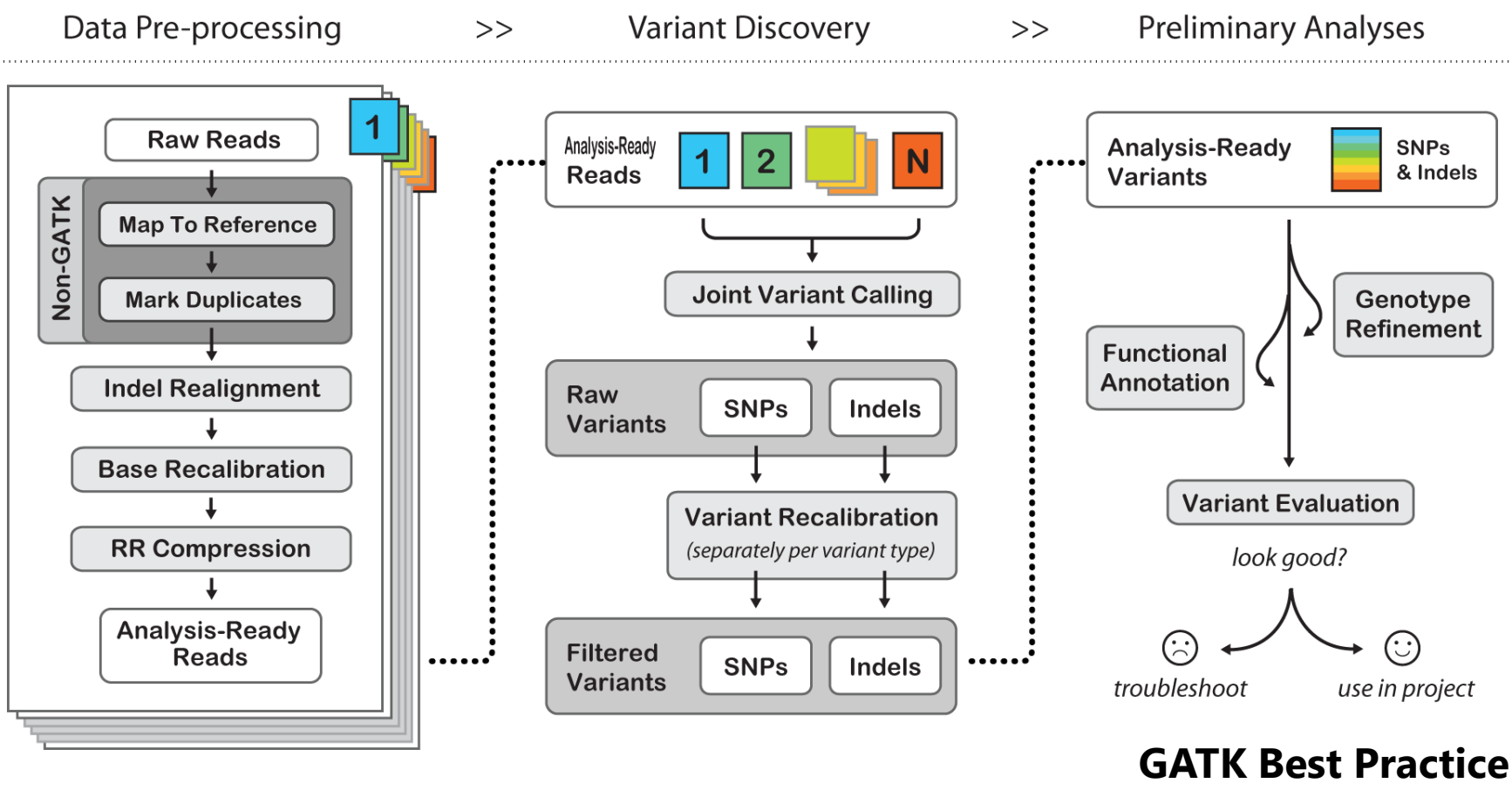
Current NHS Pipeline

Sheffield Children's NHS
NHS Foundation Trust



Typical Bioinformatics Pipeline

Sheffield Children's NHS
NHS Foundation Trust

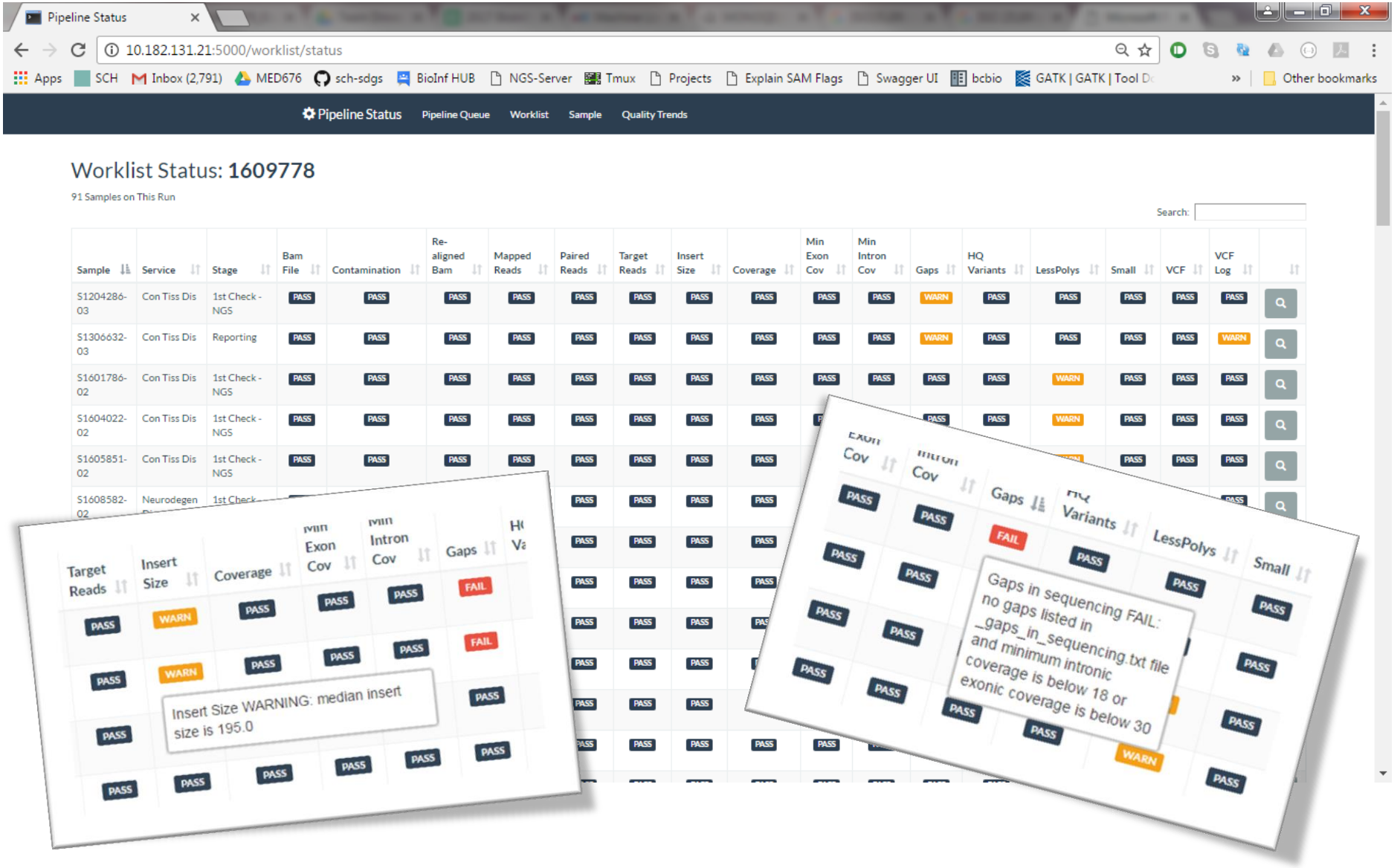


GATK Best Practice

This omits QC steps like: Low Level Read Stats, Coverage, Contamination etc…

QC

Sheffield Children's NHS
NHS Foundation Trust



Reproducibility

Sheffield Children's NHS
NHS Foundation Trust

How would you ensure that your analysis can be reproduced in say 3 years time?

Best Practice

Sheffield Children's NHS
NHS Foundation Trust

Association for Clinical Genetic Science
Part of the British Society for Genetic Medicine

http://www.acgs.uk.com/committees/quality-committee/best-practice-guidelines/

- 2 sets of guidelines:
  - Bioinformatics
  - Next Gen Sequencing
- 2 sets of statements:
  - Shall: Requirement
  - Should: Recommendation

Bioinformatics Best Practice Guidelines

acgs
Association for Clinical Genetic Science
Part of the British Society for Genetic Medicine

Guidelines for development and validation of software, with particular focus on bioinformatics pipelines for processing NGS data.

Nicola Whiffin[1,2], Kim Brugger[3], Joo Wook Ahn[4]

[1] NIHR Cardiovascular Biomedical Research Unit, Royal Brompton & Harefield NHS Foundation Trust, London
[2] National Heart and Lung Institute, Imperial College, London
[3] Department of Molecular Genetics, Addenbrooke's Hospital, Cambridge
[4] Genetics Laboratories, Guy's and St Thomas' NHS Foundation Trust, London

Requirements ("Shall")

Sheffield Children's NHS
NHS Foundation Trust

- Development
  - Write codes for humans not machines
  - House style ok but should be enforced
  - Annotation in code
  - Contingency planning & other risk assessments for all software critical for diagnostic testing
  - Version control
    - Used to record all changes
    - Informative commit messages
    - Record versions of ancillary files (Ref genome etc)
    - Include releases/milestones
    - Multiuser
  - Peer-review of code

Git is a popular versioning system

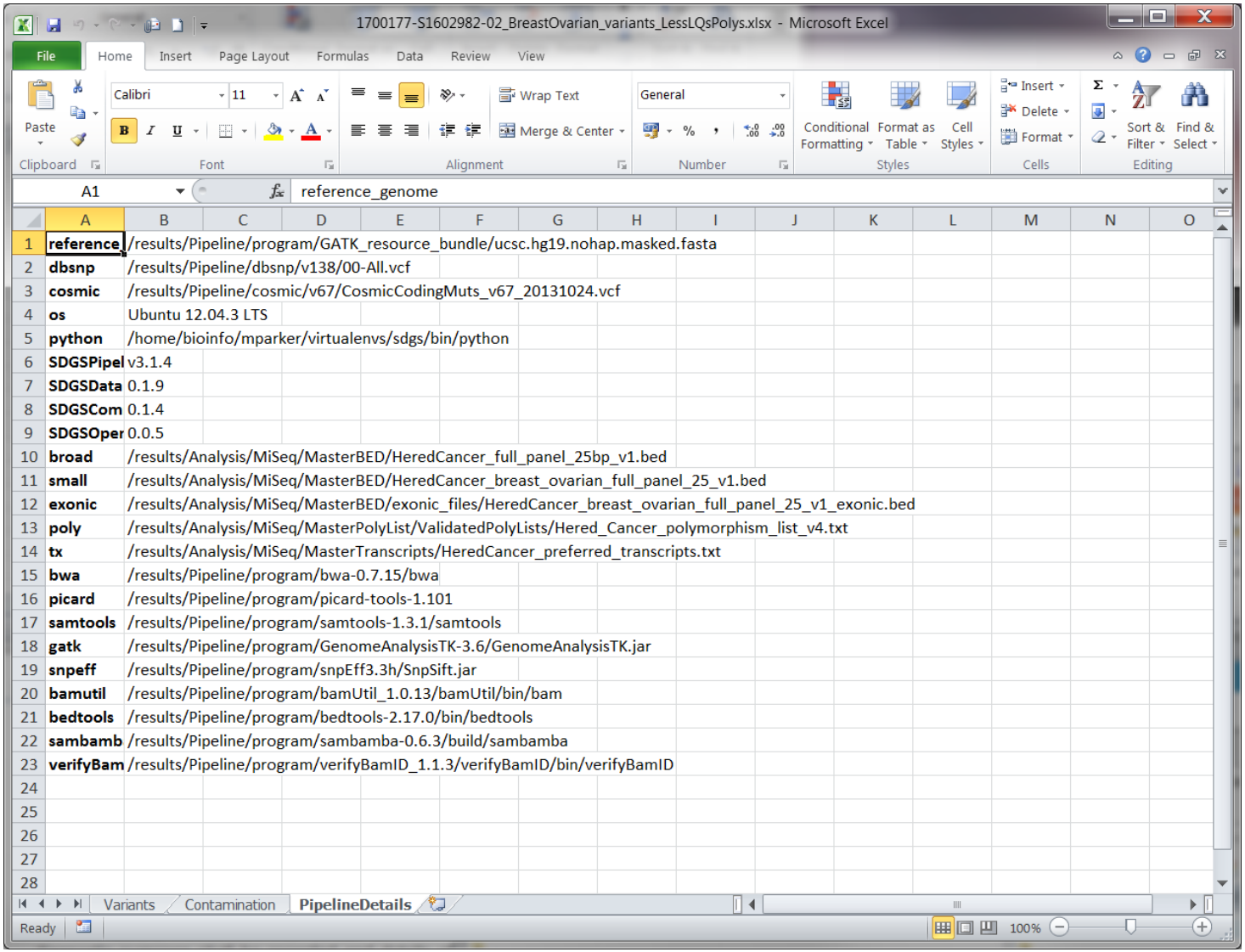
Requirements ("Shall")

Sheffield Children's NHS
NHS Foundation Trust

- External Software
  - A system for regularly checking for updates and big fixes
  - Versions of any software used for diagnostic purposes recorded
  - Key software version included on clinical report
  - Recording of ancillary files

SDGS Record

Sheffield Children's NHS
NHS Foundation Trust



1700177-S1602982-02_BreastOvarian_variants_LessLQsPolys.xlsx - Microsoft Excel

	A	B
1	**reference**	/results/Pipeline/program/GATK_resource_bundle/ucsc.hg19.nohap.masked.fasta
2	**dbsnp**	/results/Pipeline/dbsnp/v138/00-All.vcf
3	**cosmic**	/results/Pipeline/cosmic/v67/CosmicCodingMuts_v67_20131024.vcf
4	**os**	Ubuntu 12.04.3 LTS
5	**python**	/home/bioinfo/mparker/virtualenvs/sdgs/bin/python
6	**SDGSPipel**	v3.1.4
7	**SDGSData**	0.1.9
8	**SDGSCom**	0.1.4
9	**SDGSOper**	0.0.5
10	**broad**	/results/Analysis/MiSeq/MasterBED/HeredCancer_full_panel_25bp_v1.bed
11	**small**	/results/Analysis/MiSeq/MasterBED/HeredCancer_breast_ovarian_full_panel_25_v1.bed
12	**exonic**	/results/Analysis/MiSeq/MasterBED/exonic_files/HeredCancer_breast_ovarian_full_panel_25_v1_exonic.bed
13	**poly**	/results/Analysis/MiSeq/MasterPolyList/ValidatedPolyLists/Hered_Cancer_polymorphism_list_v4.txt
14	**tx**	/results/Analysis/MiSeq/MasterTranscripts/HeredCancer_preferred_transcripts.txt
15	**bwa**	/results/Pipeline/program/bwa-0.7.15/bwa
16	**picard**	/results/Pipeline/program/picard-tools-1.101
17	**samtools**	/results/Pipeline/program/samtools-1.3.1/samtools
18	**gatk**	/results/Pipeline/program/GenomeAnalysisTK-3.6/GenomeAnalysisTK.jar
19	**snpeff**	/results/Pipeline/program/snpEff3.3h/SnpSift.jar
20	**bamutil**	/results/Pipeline/program/bamUtil_1.0.13/bamUtil/bin/bam
21	**bedtools**	/results/Pipeline/program/bedtools-2.17.0/bin/bedtools
22	**sambamb**	/results/Pipeline/program/sambamba-0.6.3/build/sambamba
23	**verifyBam**	/results/Pipeline/program/verifyBamID_1.1.3/verifyBamID/bin/verifyBamID
24		
25		
26		
27		
28		

Variants | Contamination | PipelineDetails

Ready

False Positives & False Negatives

Sheffield Children's NHS
NHS Foundation Trust

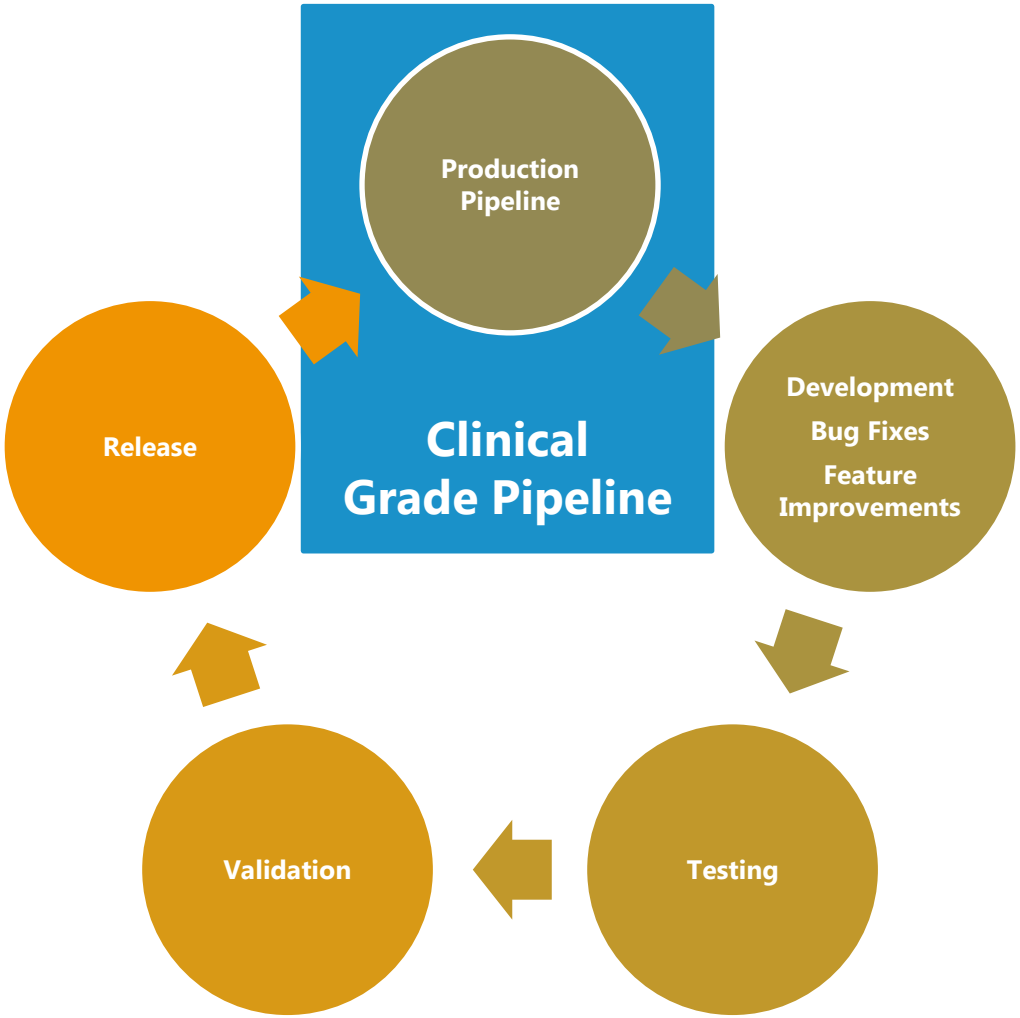
- False Positives:
  - Variants detected by NGS that are not present in the individual
    - Sequencing Artefacts
    - Poor Alignment (Pseudogenes?)
    - Poly-Tracts
- False Negatives:
  - Variants present in the patient but not present in the NGS variant calls
    - Low coverage
    - Variant caller errors
    - Poor alignment?

Requirements ("Shall")

- Validation
  - Extensive
    - Initial validation:
      - Dry – truth data
      - Determine sensitivity with sanger truth set
        - 95% >0.95 – Pipeline detects all 60/60 sanger variants with no false negatives. 300/300 gives 95% >0.99
        - Should be different individuals
    - Further validation:
      - Validation performed as above – maintain a validation dataset
    - Validation following substantive changes to a pipeline
      - Assess test dataset for relevance

Development Lifecycle

Sheffield Children's NHS
NHS Foundation Trust



Bioinformatics Guidelines

Sheffield Children's NHS
NHS Foundation Trust

Do these guidelines go far enough? What else could we do?

Sequencing Best Practice Guidelines

acgs
Association for Clinical Genetic Science
Part of the British Society for Genetic Medicine

Practice guidelines for Targeted Next Generation Sequencing Analysis and Interpretation.

Prepared and edited by Zandra Deans[1], Christopher M Watson[2], Ruth Charlton[2], Sian Ellard[3,4], Yvonne Wallis[5], Chris Mattocks[6], and Stephen Abbs[7].

1. UK NEQAS for Molecular Genetics, UK NEQAS [Edinburgh], Department of Laboratory Medicine, Royal Infirmary of Edinburgh, Edinburgh, EH16 4SA, UK
2. Yorkshire Regional DNA Laboratory, St James's University Hospital, Leeds LS9 7TF, UK.
3. Department of Molecular Genetics, Royal Devon & Exeter NHS Foundation Trust, Exeter, EX2 5DW, UK.
4. University of Exeter Medical School, Exeter, EX2 5DW, UK.
5. West Midlands Regional Genetics Laboratory, Birmingham Women's NHS Foundation Trust, Birmingham, B15 2TG, UK
6. Wessex Regional Genetics Laboratory, Salisbury District Hospital, Salisbury, SP2 8BJ, UK
7. East Anglian Medical Genetics Service, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK.

Requirements

Sheffield Children's NHS
NHS Foundation Trust

- Validation
  - Targeting method, sequencing process & data analysis
- Derive information on reproducibility and robustness

Sensitivity

- Comparison to gold standard (i.e. Sanger, SNPArrays)
- Power related to determine related to number of unique variants
- Required sensitivity depends on application
- 95% confidence that the error rate for het/hom mutation detection is no more than 5%
  - Minimum of 60 unique variants
  - 150 variants – 2%
- Use commercial standards – GIAB
- Should test on representative samples – i.e. not all cell lines

Quality Aspects

Sheffield Children's NHS
NHS Foundation Trust

- Consider difficult regions of the genome
  - Pseudogenes
  - GC/AT rich regions
  - Repetitive elements
- Barcoding – checks in place

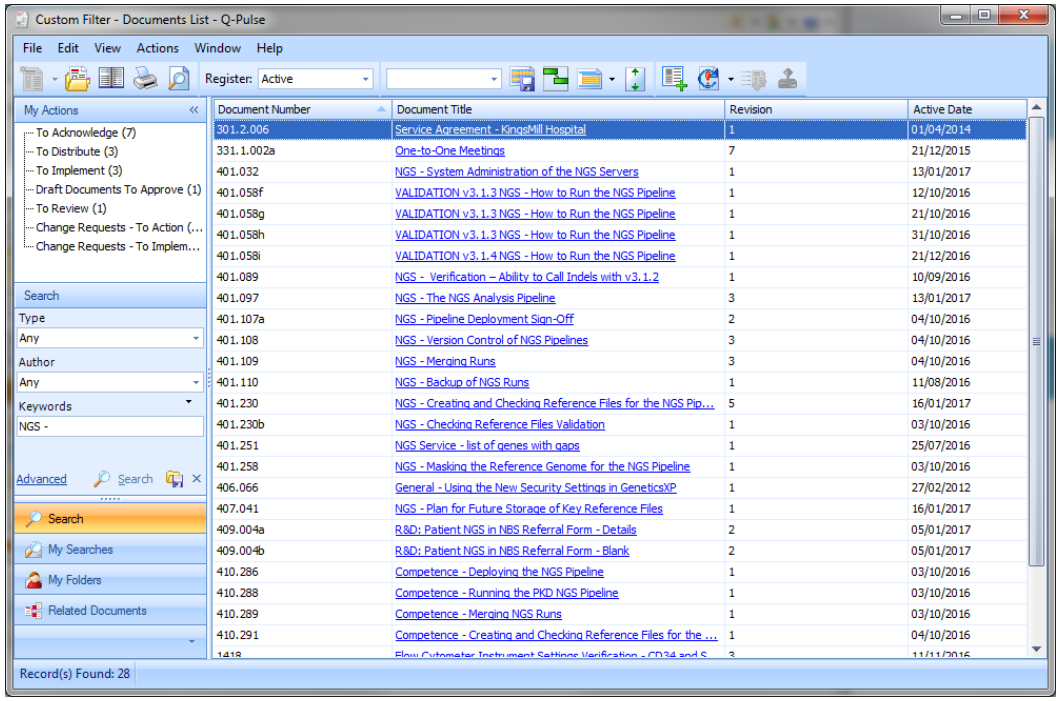
Data Analysis

- Quality measures at multiple stages of the process
- Record of data quality markers as part of audit trail:
  - Average base call quality scores
  - Mapping quality scores
  - Number of reads mapped and the percentage of target covered at the minimum coverage required
  - The alignment algorithm and alignment settings

- Analysis Pipeline to Identify Variants
  - Software for data analysis may be supplied commercially or be open source
  - Accurate versioning is essential and each software upgrade requires revalidation
- Annotation
  - Human Genome Variation Society (HGVS) recommendations
- Filtering
  - Likely mode of inheritance
  - Polys
- Data Storage
  - It is essential to store the output file from the variant annotation step

UKAS Accreditation

UKAS
UNITED KINGDOM ACCREDITATION SERVICE



Document Number	Document Title	Revision	Active Date
301.2.006	Service Agreement - KingsMill Hospital	1	01/04/2014
331.1.002a	One-to-One Meetings	7	21/12/2015
401.032	NGS - System Administration of the NGS Servers	1	13/01/2017
401.058f	VALIDATION v3.1.3 NGS - How to Run the NGS Pipeline	1	12/10/2016
401.058g	VALIDATION v3.1.3 NGS - How to Run the NGS Pipeline	1	21/10/2016
401.058h	VALIDATION v3.1.3 NGS - How to Run the NGS Pipeline	1	31/10/2016
401.058i	VALIDATION v3.1.4 NGS - How to Run the NGS Pipeline	1	21/12/2016
401.089	NGS - Verification – Ability to Call Indels with v3.1.2	1	10/09/2016
401.097	NGS - The NGS Analysis Pipeline	3	13/01/2017
401.107a	NGS - Pipeline Deployment Sign-Off	2	04/10/2016
401.108	NGS - Version Control of NGS Pipelines	3	04/10/2016
401.109	NGS - Merging Runs	3	04/10/2016
401.110	NGS - Backup of NGS Runs	1	11/08/2016
401.230	NGS - Creating and Checking Reference Files for the NGS Pip...	5	16/01/2017
401.230b	NGS - Checking Reference Files Validation	1	03/10/2016
401.251	NGS Service - list of genes with gaps	1	25/07/2016
401.258	NGS - Masking the Reference Genome for the NGS Pipeline	1	03/10/2016
406.066	General - Using the New Security Settings in GeneticsXP	1	27/02/2012
407.041	NGS - Plan for Future Storage of Key Reference Files	1	16/01/2017
409.004a	R&D: Patient NGS in NBS Referral Form - Details	2	05/01/2017
409.004b	R&D: Patient NGS in NBS Referral Form - Blank	2	05/01/2017
410.286	Competence - Deploying the NGS Pipeline	1	03/10/2016
410.288	Competence - Running the PKD NGS Pipeline	1	03/10/2016
410.289	Competence - Merging NGS Runs	1	03/10/2016
410.291	Competence - Creating and Checking Reference Files for the ...	1	04/10/2016

ISO 15189
Medical Laboratories – Requirements for Quality and Competence

ISO 15189

Sheffield Children's NHS
NHS Foundation Trust

Reporting

Sheffield Children's NHS
NHS Foundation Trust

- Negative results sometimes as important as positive ones
- Must know coverage – have we covered all regions of interest
- Gap Fills

Training

Sheffield Children's NHS
NHS Foundation Trust

- The NHS, through the Scientists Training Program (STP) is building a workforce of bioinformaticians for the future
- We have 3x trainees in Sheffield and another due to start in the new year
- Registered clinical scientists

NHS
Health Education England

National School of Healthcare Science

STP Program: Modernising Scientific Careers

Key Points

Sheffield Children's NHS
NHS Foundation Trust

- NHS already carries out significant amounts of diagnostic sequencing
- Growing bioinformatics workforce
- Genomics in healthcare is tightly regulated
- Standards and Accreditation protects patients and families from poor practice
- Bioinformatics, through a robust, version controlled pipeline, should provide high quality and relevant variants to clinical scientists