Rapid clinical whole genome sequencing: a scalable <3 day process.

SAM DELUCA, TOM HOWD, MARK FLEHARTY, FONTINA KELLEY, TIM DESMET, KATIE LARKIN, TAMARA MASON, ANDREW BERNIER, JIM MELDRIM, MAEGAN HARDEN, ALYSSA MACBETH, LEE LICHTENSTEIN, JONN SMITH, STEVEN HARRISON, HEIDI REHM, STACEY GABRIEL, NIALL LENNON

Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge MA 02141

Introduction

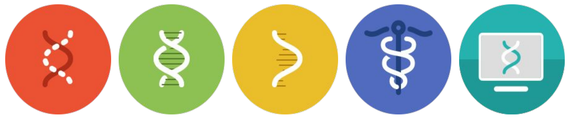
As precision medicine moves from the realm of clinical utility studies towards standard of care, there is a pressing need to decrease turnaround times for WGS data generation as part of an end-to-end clinical workflow. In the last year, we have validated and launched a clinical WGS product within the industry standard of 28 calendar days. Here, we present how we have since built a reproducible workflow around this clinical grade, gold standard PCR-Free genome process that allows for sample intake to return of data in < 3 days.

Rapid Genome Pilot Design and Goals

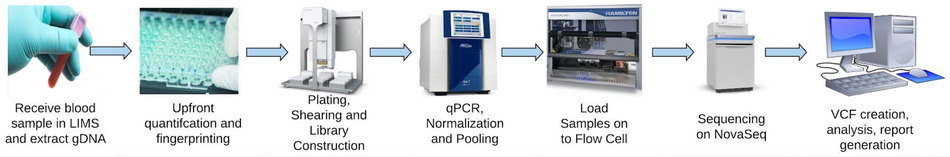
We set out to demonstrate that we could modify our high-throughput PCR-free clinical genome sequencing operation to minimize sample turnaround time starting from whole blood tubes and ending with a preliminary report across a pre-defined set of genes.

A total of six samples were run in two batches starting on consecutive days. Three of the samples started from whole blood tubes, three from genomic DNA. The samples included NA12878, which has a gold-standard curated reference dataset maintained by the National Institute of Standards and Technology*; and five samples that had been previously characterized and identified as having known pathogenic or likely pathogenic variants within the ACMG59 genes (n=3), or known to be negative for pathogenic or likely pathogenic variants in those genes (n=2).

*https://www.nist.gov/programs-projects/genome-bottle



Creating an automated and scalable workflow



								Total Time
Standard	1 Day	1 Day	2 Days	1 Day	1 Day	3 Days	5-8Days	**15-20 Days**
Run 1	47 min	74 min	273 min	295 min	58 min	23.5 hrs	180 min	**57.5 hrs**
Run 2	45 min	68 min	230 min	292 min	52 min	23.5 hrs	105 min	**37.5 hrs**

Workflow differences allowing for a rapid genome

	Lab Batch Size	Sequencing Flow Cell	Analysis Pipeline	Variant Annotation and Reporting
Standard Workflow	96 samples	NovaSeq S4 2x151	Picard and GATK	Funcotator WDL + hand off to partner lab
Rapid Workflow	1-4 samples	NovaSeq S1 2x151	Edico-Dragen	Funcotator WDL + in-house review and interpretation

Workflow Modifications

To meet the turnaround time targets we made the following modifications to our standard workflow:

- Reduced lab batch size to three samples, plus positive control.
- Eliminated sample queues at each step to reduce wait time
- Streamlined processing by reducing delays at handoff points (i.e. between extraction process, DNA QC, library construction, and loading sequencers).
- Utilized the S1 flowcell:
  - generating between 425-475 usable GB.
  - cut run time from 48 hours to 23 hours.
- Aligned reads and called variants using Edico Dragen.

	Edico Dragen pipeline	Broad Current Clinical Pipeline
bcl2fastq conversion	30 minutes (x3 samples)	4-6 days
Map, align, and mark duplicates	16 minutes/ sample	
Variant Calling	17 minutes/ sample	2-4 days
Total runtime	**63 minutes**	**6-10 days**

Analysis of Clinically Relevant Results

Following variant calling on the Dragen server, a Broad developed tool, termed Funcotator was used to annotate variants. A second in-house developed filtration script was used to filter to variants within the ACMG59 genes that are either predicted to cause loss of function or are annotated as pathogenic or likely pathogenic in internal data sources or ClinVar with a max minor allele frequency of ≤5% in ExAC. Variants were reviewed, and a preliminary report created by a certified clinical geneticist. Concordance with previously known (and orthogonally confirmed) clinically meaningful variants was evaluated. The table below summarizes these results.

Sample	Known Disease-Causing Variants in ACMG59	Rapid Genome Results
Patient Sample 1	*SCN5A* c.3573G>A (p.W1191*) Likely pathogenic for Brugada syndrome	Concordant - Disease causing variants identified
Patient Sample 2	*PCSK9* c.94G>A (p.E32K) Pathogenic for familial hypercholesterolemia	
Patient Sample 3	*MUTYH* c.536A>G(;)1187G>A p.Y179C(;)G396D Pathogenic for MYH-associated polyposis	
Patient Sample 4	None	Concordant - No disease causing variants in ACMG59 genes identified
Patient Sample 5	None	
Patient Sample 6 (NA12878)	None	

Pipeline Comparisons

All 6 samples that were run in our pilot were processed through the Edico Dragen pipeline, as well as subsequently Broad's Picard and GATK pipeline. Each sample reached approximately 50X mean coverage and satisfied all of our clinical deliverables.

Additionally, sensitivity detected for both SNV and InDels were passing our clinical thresholds. For the NA12878 sample, analytical sensitivity for SNVs were >98% and InDels analytical sensitivity were >98% when compared to NIST NA12878.

	Mean Coverage*		Marked Duplicates (%)		Mean Insert Size	
	Picard + GATK	Dragen	Picard + GATK	Dragen	Picard + GATK	Dragen
Sample 1	52.24	50.86	2.3	1.97	323	315
Sample 2	47.93	44.76	2.7	1.87	335	325
Sample 3	44.65	41.98	2.1	1.68	332	321
Sample 4	51.93	48.16	2.8	2.04	344	338
Sample 5	55.87	52.41	2.5	2.04	374	368
Sample 6	52.22	48.16	2.5	2.04	344	338

Alignment performed on hg19 with Dragen pipeline; alignment performed on hg38 for Picard and GATK pipeline for initial coverage metrics, and subsequently re-aligned to hg19 for variant calling. Differences in mean coverages are attributed to subtle differences in alignment and measurement methods yet to be fully investigated.

Conclusion

The focus of the Broad's whole genome group over the last few years has been on quality at scale, while maintaining an industry standard of < 6 week turn around time. This pilot demonstrated that for high impact clinical applications we could operate at an extremely rapid turnaround time with no loss of quality. In as little as 38 hours we can go from a blood sample to calling and reporting variants identified in a ~35x genome.

We look forward to participating in studies to establish the clinical utility of such a rapid genome product.

Acknowledgments

Data used in this poster was generated at the Broad Institute, for more information please visit: http://genomics.broadinstitute.org/