Scaling across orders of magnitude: Building cost-conscious automation for rapid growth

Justin Lock, Hoai Nguyen, Abdimalik Khalif

color

Summary

Color Genomics is disrupting the inherited disease testing marketplace by providing high quality and physician-ordered germline hereditary cancer risk testing for $249. Building automation solutions, a requirement for sustained high-quality testing that scales vertically and is also flexible to sample volume fluctuations, is a significant challenge not addressed by off-the-shelf next-generation sequencing procedures and automation products. To this end, Color Genomics has developed custom DNA purification, library preparation, target enrichment and sequencing methods specifically designed to utilize the strengths and mitigate the weakness of cutting edge next generation sequencing chemistries and commercial automation solutions. Specifically, Color Genomics has developed automation tools to simultaneously extract DNA from a variety of samples types, robustly prepare and capture libraries from DNA quickly and efficiently via rapid enzymatic fragmentation-based library preparation and end-to-end automated, single-tube capture and hybridization procedures. Furthermore, batch size flexibility was achieved through custom designed and 3D printed hardware, allowing placement of commercial reagent containers directly on robotic decks, eliminating dead volume concerns associated with variably sized batches and, more generally, streamlining robotic deck setup. By combining flexibly-sized, rapidly processed and conveniently timed wet-lab automation methods and tools, Color Genomics can provide vertically scalable, low cost and high quality molecular diagnostic products in a rapidly evolving disruptive capacity.

Introduction

Next generation sequencing (NGS) is a powerful tool that is becoming the standard of care for genomic healthcare (Richards et. al. 2015). However, genomic healthcare products are traditionally inaccessible due, individually or cumulatively, to high out-of-pocket costs (Kausmeyer et al. 2006), difficulties with obtaining samples from populations in rural communities (Abraham et al., 2012) and a general lack of knowledge surrounding the utility and availability of genomic products. Commercially available reagents and automation solutions developed for NGS methods have been designed to provide high-quality and robust results on the scale of individual academic or hospital labs, the most commonly sized lab offering genomic healthcare products. Few reagents and automation solutions exist that are designed for use in very high-throughput, quality focused and end-to-end automated labs designed to perform NGS based tests at population scales. Therefore, in order for Color Genomics to democratize genomic healthcare products and offer affordable and accessible NGS based products, we needed to source reagents and build automation solutions that maintain or increase quality, while at the same time achieving levels of scaling and end-to-end automation not available via off-the-shelf reagents or automation products. To overcome the challenges of providing population scale genomics, Color Genomics has developed custom chemistries, automation procedures and processes to meet throughput and scalability requirements, while at the same time ensuring robust and consistent assay quality. Here, we will focus on describing the procedure and lessons learned from two process common to all NGS labs, DNA purification and library preparation.

Conclusion

There are significant challenges associated with building cost-conscious, flexible and scalable NGS solutions that generate robust and high-quality results. Here, we have described the approaches used by Color Genomics to address common issues that arise during DNA purification and library preparation. By implementing a universal extraction chemistry that utilizes Omega BioTek's HDQ chemistry combined with on-the-fly generated labware sequences, Color Genomics developed a universal DNA purification procedure to routinely extract high quality genomic DNA from various collection devices, specimen types, and preservatives. Issues that arise for many labs during library preparation were successfully addressed in a three point approach. First, physical and software based procedures were deployed to eliminate sample swap and greatly reduce batch level contamination. Second, the high speed and highly efficient Kapa HyperPlus Library preparation procedure was end-to-end automated in order to limit problems that arise due to incomplete DNA fragmentation and low efficiency DNA modification steps. Lastly, fine grain sequencing results were visualized to allow rapid diagnosis and resolution of issues that might arise during the library preparation procedure. Combined, these approaches have allowed Color Genomics to develop a highly efficient, flexible and scaleable NGS workflow to drive further democratization of genomic healthcare.

References

Richards, Sue, Nazneen Aziz, Sherri Bale, David Bick, Soma Das, Julie Gastier-Foster, Wayne W. Grody, et al. "Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology." *Genetics in Medicine* 17, no. 5 (May 2015): 405–23. doi:10.1038/gim.2015.30.

Kausmeyer, Dana T., Eugene J. Lengerich, Brenda C. Kluhsman, Dorothy Morrone, Gregory R. Harper, and Maria J. Baker. "A Survey of Patients' Experiences with the Cancer Genetic Counseling Process: Recommendations for Cancer Genetics Programs." *Journal of Genetic Counseling* 15, no. 6 (December 2006): 409–31. doi:10.1007/s10897-006-9039-2.

Abraham, Jean E., Mel J. Maranian, Inmaculada Spiteri, Roslin Russell, Susan Ingle, Craig Luccarini, Helena M. Earl, Paul P. D. Pharoah, Alison M. Dunning, and Carlos Caldas. "Saliva Samples Are a Viable Alternative to Blood Samples as a Source of DNA for High Throughput Genotyping." *BMC Medical Genomics* 5 (May 30, 2012): 19. doi:10.1186/1755-8794-5-19.

Methods

DNA Purification

The two major challenges associated with the development of a universal DNA purification procedure, the removal of specimen specific contaminants and physical differences in settling rates for lysis and unlased input materials were addressed using Omega Bio-tek's Mag-Bind® Blood & Tissue DNA HDQ implemented on the Hamilton Star platform:

1. Saliva samples collected in Oragene DX collection devices are spun at 2000 g for two minutes to pellet particulates before being transferred, carefully as to not disturb the pellet, to the Hamilton deck for initiation of the extraction procedure.
2. The linear barcode label attached to input samples is scanned by the Hamilton Star's Autoloader during deck setup and, based on the barcode, the sample type is identified and the appropriate labware parameters are established to allow for labware specific sample handling:
   a. Saliva: An aliquot (300 µL) of each saliva sample is carefully aspirated from 1 mm below the meniscus of provided samples using slow aspiration speeds (5 µL/s) and transferred to the lysis plate. This process prevents disturbing pelleted materials.
   b. Peripheral Blood: Samples are pipette mixed (800 µL or 80 % of liquid volume, whichever is smallest) to resuspend settled cells before an aliquot is transferred to the lysis plate.

Library Preparation

The three major sources of library preparation issues were concurrently addressed using physical, software and process optimizations.

1. Sample Swaps: Eliminating the risk and ensuring, nonetheless, that sample swaps are detected prior to data finalization was addressed using five different approaches:
   a. Sample and index plates are physically barcoded on only one side and never manually manipulated or handled, single-use and are hermetically sealed. Additionally, the indexes used in a given batch are randomized (768 unique indexes with >8bp edit distance).
   b. At points in the lab process where samples are in individual Micronics tubes, each sample is barcoded and this barcode is scanned by the Hamilton prior to any transfer or pipetting steps to ensure that the sample is in the correct well position and on the correct plate.
   c. Three control samples (two characterized cell lines and one no-template control) are randomly assigned to different well locations in every library preparation batch.
   d. Various post-pipeline quality control checks are performed on all samples in every batch and include the identification of low fraction alleles indicative of low level contamination, kin-ship type analysis within and between recent lab batches and a gender check.
2. Incomplete fragmentation: Ensuring and monitoring the success of fragmentation was addressed using two approaches:
   a. Only samples with DNA quality and purity standards that meet purpose designed thresholds are input into library preparation (hybrid QC metric of 1.3 hybrid QC units) as defined below:

$$\textit{Hybrid QC} = (A260{:}A280/1.8) + (A260{:}A230/1.2)$$

   b. Fragmentation success is assessed routinely at completion of sequencing to monitor and ensure assay drift is not occurring. These measures include distribution of yield, insert-sizes assessment of GC/AT bias and distribution of read counts.
3. Low efficiency: Ensuring library preparation efficiency was addressed using three approaches:
   a. Kapa BioSystem's HyperPlus Library Preparation system was automated on the Hamilton Star platform. The chemistry is performed in a single tube and has includes two bead based purification steps. All incubation steps are performed on deck using an oil overlay to prevent evaporation.
   b. Reagent master mixes are prepared entirely on deck by placing stock reagents into a custom 3D printed holder. Master mix aliquots dispensed into 384 well plates prior to being stamped into sample wells to speed up reagent mixing steps and reduce required dead volumes.

Results

DNA Purification

Extracting high-quality DNA and removal of contaminants, or the mitigation of their effects in downstream assays, is a requirement for the development of a universal DNA purification procedure. In addition, a cell lysis procedure that can take as input material already lysed and also fully intact cells without compromising DNA yield, purity or molecular weight of extracted DNA was a required. Both of these issues were overcome by the addition of physical and chemical sample handling procedures. The addition of a pre-extraction centrifugation step significantly reduced contaminants from saliva samples (Figure 1). The sample specific processing based on the scanned linear barcode of input samples results in low levels of variability in the purity and concentration of different specimen types extracted simultaneously using a universal extraction procedure (Figure 2).

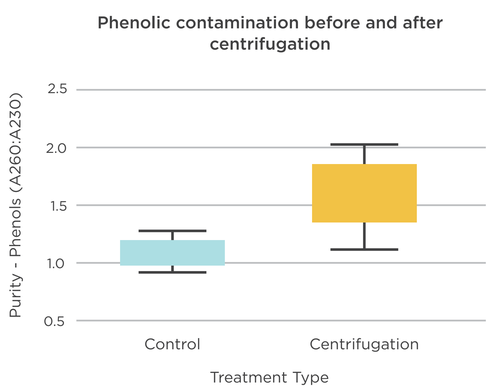

Figure 1: The results of a study to determine the impact of a brief low force centrifugation of Oragene Dx saliva samples prior extraction on level of phenolic contamination. Centrifugation has a significant positive impact on purity (Paired-Samples t-Test$_{22}$ = 8.353, p < 0.001).

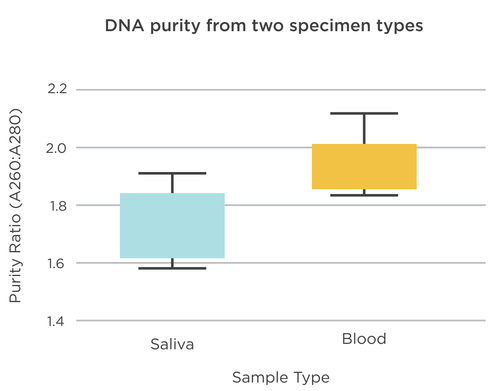

Figure 2: The purity of DNA extracted from two specimen types using Omega Bio-tek's Mag-Bind® Blood & Tissue DNA HDQ chemistry and automated in parallel using the Microlab STAR platform. The method extracted DNA from 300 µL of starting material (saliva, n = 500 or peripheral blood, n = 95) and eluted purified nucleic acids in 60 µL of 10 mM Tris-HCl before measurement of DNA purity via spectrophotometry (A260:A280) using the Trinean DropSense96 instrument. Whiskers show the 10th and 90th percentiles.

Library Preparation

Sample swaps caused by manual sample processing are eliminated via removal of all manual handling steps and complete automation of sample processing, including the addition of barcoded and orientation specific labware and consumables and the placement of reagent containers directly onto the Hamilton Star deck. Cross and background contamination are monitored and assessed in every batch (Figure 3). Fragmentation success and consistency is robust within and between specimen types (Figure 4). If performance issues do arise that impact assay efficiency during library preparation, heat map style data visualizations allow simple troubleshooting and quick issue resolution (Figure 5).

	1	2	3	4	5	6	7	8
A	NTC	0.13%	0.02%	0.09%	0.26%	0.08%	0.04%	0.03%
B	0.03%	NTC	0.07%	0%	0.07%	0.04%	0.06%	0.06%
C	0.08%	0.14%	NTC	0.05%	0.11%	0%	0.03%	0.04%
D	0.11%	0.12%	0.09%	NTC	0.03%	0.02%	0%	0.02%
E	0.06%	0.06%	0.09%	0.07%	NTC	0.03%	0.07%	0.03%
F	0.06%	0%	0.01%	0.06%	0.03%	NTC	0.10%	0.03%
G	0.03%	0%	0%	0.03%	0.04%	0.04%	NTC	0.10%
H	0.03%	0.04%	0.02%	0.07%	0%	0.06%	0.06%	NTC

Figure 3: The levels of contamination detected across a sample plate after library preparation, enrichment and sequencing of 30 inherited cancer genes.

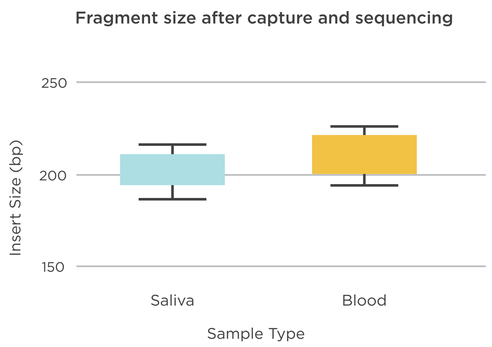

Figure 4: The insert size of library fragments mapped to the human reference genome after library preparation (Using Kapa Biosystems' HyperPlus Library Prep Kit) and capture (Agilent's SureSelect XT Kit) using input gDNA from two two specimen types. Specimens (saliva, n = 500, peripheral blood, n = 95) were extracted with Omega Bio-tek's Mag-Bind® Blood & Tissue DNA HDQ chemistry automated in parallel on the Microlab STAR platform. Generated NGS data was aligned with BWA to the GRCh37 (hg19). Whiskers show the 10th and 90th percentiles.

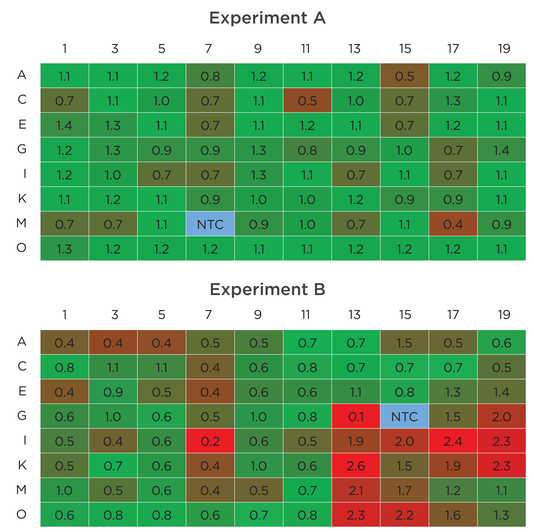

Figure 5: The distribution of insert sizes relative to the batch's median size for two different experiments. Experiment "A" (top) shows successful fragmentation across the entire sample plate. Experiment "B" (bottom) shows poor fragmentation and therefore larger insert sizes in the bottom right quadrant of the plate.