Enabling population-scale PCR-free whole genome sequencing

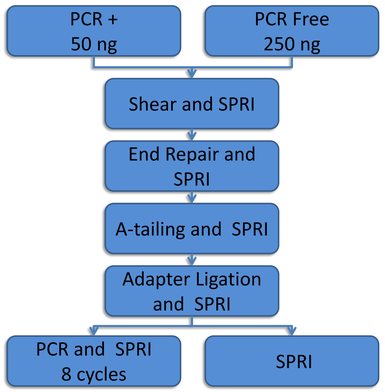
Danielle Dionne, Ryan Hegarty, Justin Abreu, Matthew Coole, Laurie Holmes, Matthew Lee, Sheli McDonough, Alvin Tam, Tim DeSmet, Bertrand Haas, Yossi Farjoun, Danielle Perrin, Niall Lennon, Sheila Fisher, Stacey Gabriel

Genomics Platform, Broad Institute, 320 Charles Street Cambridge, MA

PCR-free human whole genome libraries sequenced on the Illumina HiSeq X are the current NGS gold standard for population-scale genomics, Mendelian disease gene discovery, and understanding the genetic components of common diseases. Within the last year we have implemented a high throughput whole genome pipeline which:

- Uses a streamlined and unified workflow
- Allows for a reduction in gDNA input
- Allows for the creation of more complex sequencing pools and more even library representation
- Increases per lane yield (Gb) on the HiseqX

Increasing Scale with Unified Automated Workflow



- Single, simplified workflow for both PCR free and PCR plus protocols
- Allows a single full time employee to create up to 192 whole genome libraries per day
- Protocol is dictated by amount of available input material
  - PCR Free input: 250 ng
  - PCR+ input: 50 ng

Figure 1: Process map. Single unified library preparation for all WGS samples

Reducing Input into WGS Library Preparation

- PCR Free library prep results in improved coverage across the genome and reduction in base specific biases that are attributed to DNA polymerases
- Optimized PCR Free library preparation allows for a reduction of gDNA input while maintaining quality and sufficient yield

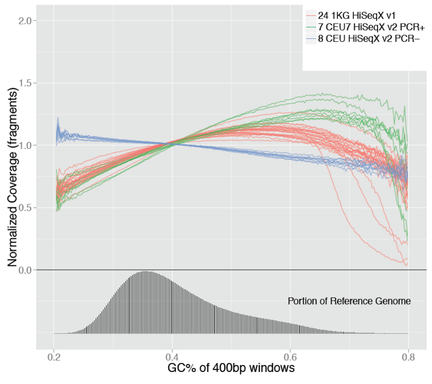


Figure 2: GC bias comparison. PCR-free WGS libraries (blue) show significantly more even coverage across the GC spectrum when compared to traditional WGS libraries (red and green).

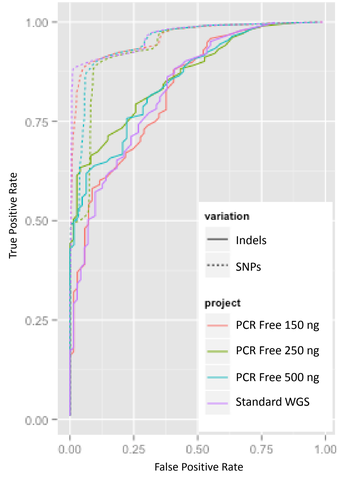


Figure 3: SNP and Indel Calling. Analysis has shown no significant difference in SNP and Indel calling between inputs of 500 ng and 250 ng gDNA into PCR Free library prep

Increasing Sequencing Yield

Increased sequencing yield achieved through:

- Improvements to automated steps and liquid handling
- Reduced downtime between strip tube creation and cluster amplification
- Identification of optimal loading concentrations by sample type to increase %PF while keeping optical duplication levels low

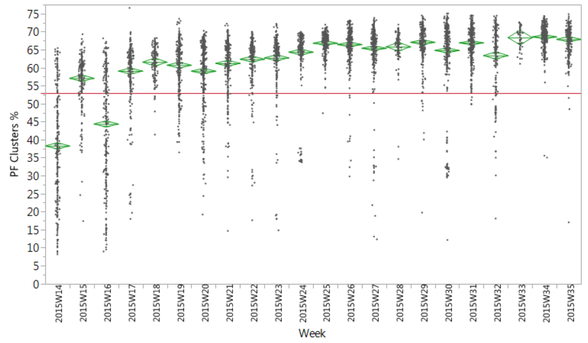


Figure 4: %PF Clusters. Process improvements increase %PF over time while decreasing variability

- Optical duplication by pad hopping is specific to the arrayed flowcell technologies. With insufficient DNA loading, a strand of DNA will "hop" from an already seeded well to a neighboring one causing identical clusters.

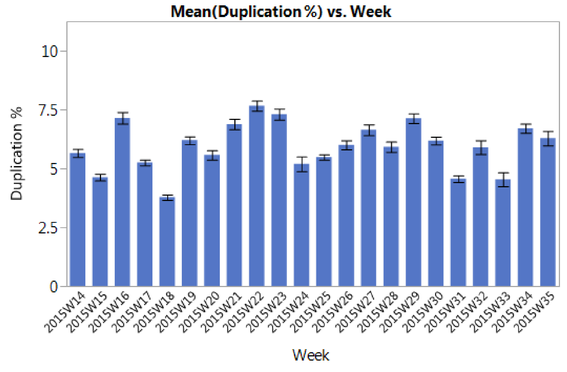


Figure 5: Duplication %. As %PF increased there was no change in overall duplication and more importantly optical duplication

Bulk denature of all lanes of a single pool allows for:

- Reduction in variability lane to lane
- Increased loading concentration resulting in a reduction of optical duplication %

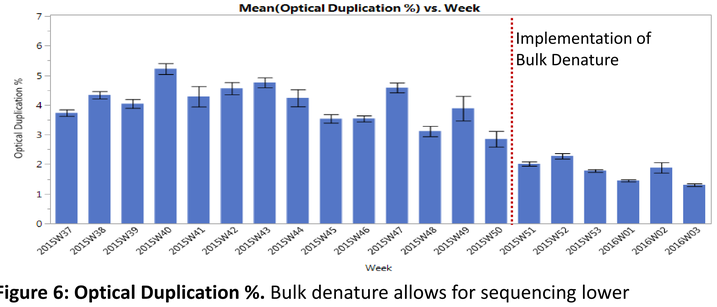


Figure 6: Optical Duplication %. Bulk denature allows for sequencing lower concentration pools at a higher loading concentration leading to an overall drop in optical duplication %.

Increasing Pooling Plexity

- Ability to pool up to 24 unique whole genome libraries prior to sequencing
- Leverage single lane of HiseqX sequencing to assess the evenness of library representation within pools
- Ability to correct for underrepresented libraries within pools to ensure desired coverage is met while minimizing the amount of sequencing necessary

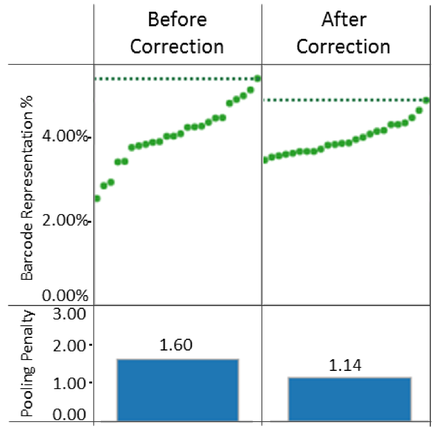
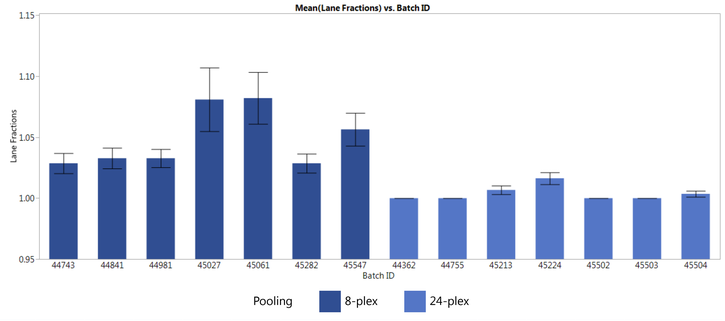


Figure 7: Pool Correction. Pool Correction strategy improves library representation from original pool (left) to adjusted pool (right).

Pool Correction results in more even pools which:

- Increases number of samples that hit coverage goal on first pass
- Leads to shorter turn around time for projects
- Reduces overall number of lanes of sequencing required per batch



Batch ID	44743	44841	44981	45027	45061	45282	45547	44362	44755	45213	45224	45502	45503	45504
1st Pass success	86.4	84.2	80.6	81.1	80.0	85.1	79.8	100	100	95.6	89.5	100	100	97.9

Figure 8: Pool Correction of 24 Plex Pools. Corrected pools result in more samples hitting coverage on first pass and fewer overall lanes being sequenced. 24 plex pools are corrected, 8 plex are not.

Acknowledgements

National Human Genome Research Institute

NIH National Heart, Lung, and Blood Institute

illumina