High Throughput Sample Preparation for Genomic Studies

Shawn Levy, PhD
Faculty Investigator
HudsonAlpha Institute for Biotechnology
Huntsville, AL
slevy@hudsonalpha.org

Adjunct Associate Professor
Department of Biomedical Informatics
Vanderbilt University Medical Center
Department of Epidemiology
Department of Genetics
University of Alabama-Birmingham

LevyLab
@genomicslab
slevy@hudsonalpha.org
genomics@hudsonalpha.org
www.hudsonalpha.org/gsl/
A fee-for-service core facility supporting both internal projects as well as projects from around the globe.

Founded in August 2009, over 45,000 samples in more than 500 projects processed since January 2010.

Certified Service Provider for Illumina Sequencing and Array Technologies. Certifications for Nimblegen, Agilent, Ion Torrent are in process.
The success of any technology center relies on the ability to appropriately quality control and analyze the data being produced in that center.

A dedicated, experienced and efficient analysis group is the center of our capabilities.

Software tools produced in-house, from open source efforts and commercial are all used.
Summary of Operations

Similar to analysis and informatic capabilities, the cutting-edge nature of most genomic technologies means there is often room for innovation and optimization.

We actively test and develop protocols to enable novel experimental designs, extend capabilities and gain efficiency or cost savings.
Summary of Operations

Basic research interests leverage the previously described capabilities and resources to explore specific areas of interest or collaborative studies.

- Whole Genome Sequencing in Bipolar Disorder
- Genetics of Sporadic Schizophrenia
- Variant Discovery in novel and rare diseases.
- Microbiome diversity and its relationship to immune system diversity.

Ultra-dense barcoding and indexing of samples.
An example run batch:

7 HiSeqs and 4 GAs running a total of 13 Flowcells

104 Pooled Sets

>150 Individual Libraries

3 Hudson Alpha Investigators (Levy, Myers, Barsh)

>15 Investigators from at least 3 different countries

>15 different species including Dropshila, Mouse, Zebrafish, Snakes, Wild Cats, Human, Catfish, Dog, mosquito

>8 Methodologies including Nextera RNA prep, Agilent Exome capture, Nimblegen Solution based, Nimblegen Array based, Ovation, Illumina RNA Prep

>4 Sample types including RNAseq, ChIPseq, Exome, small RNA, whole genome
Automated Protocols

Sample Handling
Full plate normalization and cherry picking

Library preparation
Illumina TruSeq DNA
Illumina TruSeq RNA*
GSL DNA (high-input)
GSL DNA (low input)
NuGEN Ovation

Quality Control
Kappa Biosystems real-time PCR
Final Library Dilution
Final Library Pooling
Pre-cluster denaturation and dilutions

Sequence Capture
Nimblegen Post-Hyb processing**
Agilent Post-Hyb processing**

Notes:
1. Tip wash and orbital shaker are required accessories
2. Extra heating positions are not needed for library preparation

*On-deck thermocycler required
**Non-Beckman accessories needed
Automation is vital for the accurate and efficient use of sequencing workflows.

- Automation of the bead purifications can substantially increase yields and consistency.
- Automation of the library dilution and pooling greatly aids in the final quality and accuracy of the experiment.

Automation solutions should provide a substantial improvement to both the efficiency of the experiment as well as the quality of the data.
Automation is vital for the accurate and efficient use of sequencing workflows.

- Automation of the bead purifications can substantially increase yields and consistency.
- Automation of the library dilution and pooling greatly aids in the final quality and accuracy of the experiment.

Two project examples

Low-depth, whole-genome sequencing in Bipolar Disorder Type 1

Deep exome sequencing for recessive diseases and de-novo mutations
Molecular Process

- Sample QC
- Covaris Shearing
- Library Preparation

New England Biolabs
Modified Illumina Adaptors
Kapa Biosystems
Molecular Process

Sample QC

Covaris Shearing

Library Preparation

Sequence Capture

Quality Control

Sequencing

Roche Nimblegen v2 and v3
IDT Oligos
Kapa Biosystems

PicoGreen
Caliper LabChip GX
Kapa Biosystems

Illumina HiSeq 2000
PE50 or PE100
Low-Depth sequencing for Efficient Experimental Design

Method

Low-coverage sequencing: Implications for design of complex trait association studies

Yun Li,1,4 Carlo Sidore,2,3,4 Hyun Min Kang,4 Michael Boehnke,4 and Gonçalo R. Abecasis4,5

1Department of Genetics, Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina 27599-7264, USA; 2Istituto di Neurogenetica e Neurofarmacologia, CNR, Monserrato, 09042 Cagliari, Italy; 3Dipartimento di Scienze Biomediche, Università di Sassari, Sassari 07100, Italy; 4Center for Statistical Genetics, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan 48109-2029, USA
Sample Indexing and Pooling

12-plex indexing

Stage 1 (x1)

PS1

PS8

Stage 2

8 lanes per PS (8 flowcells/plate)

PS1

PS8

9 lanes of sequencing per 12-sample pool

9 flowcells/plate
Consistency in Library Preparation
Consistency in large projects

Mean Coverage

% Aligned
Low-coverage sequencing QC
Genome-scale Sequencing

- **Whole-Genome**
  
  - 3,000,000,000bp
  
  - ATGCTCGAGTACCGAT
  
  - TGCAAGCTCGATCGAT
  
  - CGATCGATCGATCCGA
  
  - TCGATTGCATGCAACG
  
  - TTGCAAGCTAGCATCG
  
  - GCTAGCTAGCCTAGCT
  
  - AGCTAGCTAGCTAGCT
  
  - 90,000,000,000 bp

- **Exome**
  
  - 60,000,000bp
  
  - ATGCTCGAGTACCGAT
  
  - TGCAAGCTCGATCGAT
  
  - CGATCGATCGATCCGA
  
  - TCGATTGCATGCAACG
  
  - TTGCAAGCTAGCATCG
  
  - GCTAGCTAGCCTAGCT
  
  - AGCTAGCTAGCTAGCT

- **Whole-Genome**
  
  - 3,000,000,000bp
  
  - ATGCTCGAGTACCGAT
  
  - TGCAAGCTCGATCGAT
  
  - CGATCGATCGATCCGA
  
  - TCGATTGCATGCAACG
  
  - TTGCAAGCTAGCATCG
  
  - GCTAGCTAGCCTAGCT
  
  - AGCTAGCTAGCTAGCT

- **Exome**
  
  - 60,000,000bp
  
  - ATGCTCGAGTACCGAT
  
  - TGCAAGCTCGATCGAT
  
  - CGATCGATCGATCCGA
  
  - TCGATTGCATGCAACG
  
  - TTGCAAGCTAGCATCG
  
  - GCTAGCTAGCCTAGCT
  
  - AGCTAGCTAGCTAGCT

- **Whole-Genome**
  
  - 3,000,000,000bp
  
  - ATGCTCGAGTACCGAT
  
  - TGCAAGCTCGATCGAT
  
  - CGATCGATCGATCCGA
  
  - TCGATTGCATGCAACG
  
  - TTGCAAGCTAGCATCG
  
  - GCTAGCTAGCCTAGCT
  
  - AGCTAGCTAGCTAGCT

- **Exome**
  
  - 60,000,000bp
  
  - ATGCTCGAGTACCGAT
  
  - TGCAAGCTCGATCGAT
  
  - CGATCGATCGATCCGA
  
  - TCGATTGCATGCAACG
  
  - TTGCAAGCTAGCATCG
  
  - GCTAGCTAGCCTAGCT
  
  - AGCTAGCTAGCTAGCT

- **Whole-Genome**
  
  - 3,000,000,000bp
  
  - ATGCTCGAGTACCGAT
  
  - TGCAAGCTCGATCGAT
  
  - CGATCGATCGATCCGA
  
  - TCGATTGCATGCAACG
  
  - TTGCAAGCTAGCATCG
  
  - GCTAGCTAGCCTAGCT
  
  - AGCTAGCTAGCTAGCT

- **Exome**
  
  - 60,000,000bp
  
  - ATGCTCGAGTACCGAT
  
  - TGCAAGCTCGATCGAT
  
  - CGATCGATCGATCCGA
  
  - TCGATTGCATGCAACG
  
  - TTGCAAGCTAGCATCG
  
  - GCTAGCTAGCCTAGCT
  
  - AGCTAGCTAGCTAGCT

- **Whole-Genome**
  
  - 3,000,000,000bp
  
  - ATGCTCGAGTACCGAT
  
  - TGCAAGCTCGATCGAT
  
  - CGATCGATCGATCCGA
  
  - TCGATTGCATGCAACG
  
  - TTGCAAGCTAGCATCG
  
  - GCTAGCTAGCCTAGCT
  
  - AGCTAGCTAGCTAGCT

- **Exome**
  
  - 60,000,000bp
  
  - ATGCTCGAGTACCGAT
  
  - TGCAAGCTCGATCGAT
  
  - CGATCGATCGATCCGA
  
  - TCGATTGCATGCAACG
  
  - TTGCAAGCTAGCATCG
  
  - GCTAGCTAGCCTAGCT
  
  - AGCTAGCTAGCTAGCT

- **Whole-Genome**
  
  - 3,000,000,000bp
  
  - ATGCTCGAGTACCGAT
  
  - TGCAAGCTCGATCGAT
  
  - CGATCGATCGATCCGA
  
  - TCGATTGCATGCAACG
  
  - TTGCAAGCTAGCATCG
  
  - GCTAGCTAGCCTAGCT
  
  - AGCTAGCTAGCTAGCT

- **Exome**
  
  - 60,000,000bp
  
  - ATGCTCGAGTACCGAT
  
  - TGCAAGCTCGATCGAT
  
  - CGATCGATCGATCCGA
  
  - TCGATTGCATGCAACG
  
  - TTGCAAGCTAGCATCG
  
  - GCTAGCTAGCCTAGCT
  
  - AGCTAGCTAGCTAGCT

- **Whole-Genome**
  
  - 3,000,000,000bp
  
  - ATGCTCGAGTACCGAT
  
  - TGCAAGCTCGATCGAT
  
  - CGATCGATCGATCCGA
  
  - TCGATTGCATGCAACG
  
  - TTGCAAGCTAGCATCG
  
  - GCTAGCTAGCCTAGCT
  
  - AGCTAGCTAGCTAGCT

- **Exome**
  
  - 60,000,000bp
  
  - ATGCTCGAGTACCGAT
  
  - TGCAAGCTCGATCGAT
  
  - CGATCGATCGATCCGA
  
  - TCGATTGCATGCAACG
  
  - TTGCAAGCTAGCATCG
  
  - GCTAGCTAGCCTAGCT
  
  - AGCTAGCTAGCTAGCT

- **Whole-Genome**
  
  - 3,000,000,000bp
  
  - ATGCTCGAGTACCGAT
  
  - TGCAAGCTCGATCGAT
  
  - CGATCGATCGATCCGA
  
  - TCGATTGCATGCAACG
  
  - TTGCAAGCTAGCATCG
  
  - GCTAGCTAGCCTAGCT
  
  - AGCTAGCTAGCTAGCT

- **Exome**
  
  - 60,000,000bp
  
  - ATGCTCGAGTACCGAT
  
  - TGCAAGCTCGATCGAT
  
  - CGATCGATCGATCCGA
  
  - TCGATTGCATGCAACG
  
  - TTGCAAGCTAGCATCG
  
  - GCTAGCTAGCCTAGCT
  
  - AGCTAGCTAGCTAGCT
Hybridization

Hybridization

POST-HYB WASH

POST-HYB WASH

Capture Specificity

Temperature and mixing is VITAL for efficient sequencing performance.

Post-Hyb automation/high-throughput is not straightforward.

Plate

Tube

- On-target
- +/- 100bp
- +/- 200bp
- Non-specific

0% 20% 40% 60% 80% 100%
Capture and indexing summary

Library Prep → Capture → Sequencing

Library Prep → Capture → Sequencing

Library Prep → Capture → Sequencing
Pre-Capture Indexing

Percentage of Reads Mapped to Target (+/- 200 bp)

One Library per Enrichment
Pooled After Enrichment

Two Libraries per Enrichment
Pooled Before Enrichment
One Library per Enrichment
Pooled After Enrichment

Two Libraries per Enrichment
Pooled Before Enrichment
The Highlights

• 45 sample rapid-turn around project
  – 96% of clusters passing filter
  – 92% of reads >Q30
  – Mean quality score of 36
Cluster Variability Before Comprehensive Normalization
Caliper GX Library Trace
Cluster Variability After Comprehensive Normalization
Representative Exome Sequencing with automation
Method-Specific Variation in DNA quantitation

Calculated nM of "10 nM" Dilution

Nextera Batch 7.16.11
Nextera Batch 8.3.11
TruSeq Batch 5.28.11
TruSeq Batch 6.21.11
Exome Batch 6.21.11
Exome Batch 7.11.11

n = 51  n = 18  n = 88  n = 95  n = 40  n = 28
## PCR to validate robot accuracy

<table>
<thead>
<tr>
<th>Sample Description</th>
<th>Qty</th>
<th>size</th>
<th>Calc. nM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test-4-Replicate A-1</td>
<td>2.688074</td>
<td>296</td>
<td>39.50</td>
</tr>
<tr>
<td>Test-4-Replicate A-2</td>
<td>2.552462</td>
<td>296</td>
<td>37.51</td>
</tr>
<tr>
<td>Test-4-Replicate A-3</td>
<td>2.933796</td>
<td>296</td>
<td>43.11</td>
</tr>
<tr>
<td>Test-4-Replicate B-1</td>
<td>3.122124</td>
<td>296</td>
<td>45.88</td>
</tr>
<tr>
<td>Test-4-Replicate B-2</td>
<td>2.757545</td>
<td>296</td>
<td>40.52</td>
</tr>
<tr>
<td>Test-4-Replicate B-3</td>
<td>3.383912</td>
<td>296</td>
<td>49.73</td>
</tr>
<tr>
<td>Test-4-Replicate C-1</td>
<td>2.771373</td>
<td>296</td>
<td>40.73</td>
</tr>
<tr>
<td>Test-4-Replicate C-2</td>
<td>2.451192</td>
<td>296</td>
<td>36.02</td>
</tr>
<tr>
<td>Test-4-Replicate C-3</td>
<td>2.825558</td>
<td>296</td>
<td>41.52</td>
</tr>
</tbody>
</table>
Deck Layout for Nimblegen Method
The left pod should have no tips loaded. The right pod should have no tips loaded.

Does the Biomek® Software deck match the above layout, including the labware and their locations?

If yes, choose OK to continue the method. If no, choose Abort to stop the method.

WARNING: This method was configured with Wash_Station_with_Standard_Shaker but Wash_Station_Orbital_SPelt is the current deck.
# Illumina Library Construction Reagent Calculator

## Ampure Reagent Reservoir

<table>
<thead>
<tr>
<th>Reservoir 1:</th>
<th>Reservoir 2:</th>
<th>Reservoir 3:</th>
</tr>
</thead>
<tbody>
<tr>
<td>64552.8 µl of AMPureXP</td>
<td>27804 µl of Sizing Solution</td>
<td>38940 µl of Elution Buffer</td>
</tr>
</tbody>
</table>

## 24 Position Tube Block

<table>
<thead>
<tr>
<th>1228 µl of End Repair MM</th>
<th>1228 µl of A-Tailing MM</th>
<th>1111 µl of Ligation MM</th>
<th>1111 µl of Ligation MM</th>
<th>1470 µl of PCR MM</th>
<th>Empty</th>
</tr>
</thead>
<tbody>
<tr>
<td>1228 µl of End Repair MM</td>
<td>1228 µl of A-Tailing MM</td>
<td>1111 µl of Ligation MM</td>
<td>1111 µl of Ligation MM</td>
<td>1470 µl of PCR MM</td>
<td>Empty</td>
</tr>
<tr>
<td>1228 µl of End Repair MM</td>
<td>1228 µl of A-Tailing MM</td>
<td>1111 µl of Ligation MM</td>
<td>1111 µl of Ligation MM</td>
<td>Empty</td>
<td>Empty</td>
</tr>
<tr>
<td>1228 µl of End Repair MM</td>
<td>1228 µl of A-Tailing MM</td>
<td>1111 µl of Ligation MM</td>
<td>1111 µl of Ligation MM</td>
<td>Empty</td>
<td>Empty</td>
</tr>
</tbody>
</table>

### To Make End Repair Master Mix:
- 2475 µl of End Repair Buffer
- 2475 µl of End Repair Enzyme

### To Make A-Tailing Master Mix:
- 2475 µl of A-Tailing Buffer
- 2475 µl of A-Tailing Enzyme

### To Make Ligation Master Mix:
- 6532.5 µl of Ligase Buffer
- 2512.5 µl of Ligase Enzyme

### To Make PCR Master Mix:
- 2450 µl of PCR Master Mix
- 490 µl of Primer Cocktail

[OK]
### Illumina Indexing

<table>
<thead>
<tr>
<th></th>
<th>Illumina_1</th>
<th>Illumina_2</th>
<th>Illumina_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legible Reads</td>
<td>98.95%</td>
<td>99.13%</td>
<td>99.09%</td>
</tr>
<tr>
<td>Perfect</td>
<td>96.22%</td>
<td>96.49%</td>
<td>96.39%</td>
</tr>
<tr>
<td>Auto-corrected</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Heuristic</td>
<td>2.73%</td>
<td>2.64%</td>
<td>2.70%</td>
</tr>
<tr>
<td>Illegible</td>
<td>1.05%</td>
<td>0.87%</td>
<td>0.91%</td>
</tr>
<tr>
<td>Total # of Reads</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Paired-end adjustments:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rescued Reads</td>
<td>0.74%</td>
<td>0.70%</td>
<td>0.73%</td>
</tr>
<tr>
<td>Conflicting</td>
<td>0.08%</td>
<td>0.08%</td>
<td>0.09%</td>
</tr>
<tr>
<td>AAGACCT</td>
<td>6.15%</td>
<td>7.29%</td>
<td>8.40%</td>
</tr>
<tr>
<td>ACACCTCT</td>
<td>7.73%</td>
<td>7.47%</td>
<td>8.76%</td>
</tr>
<tr>
<td>ACTTAGG</td>
<td>8.72%</td>
<td>7.85%</td>
<td>7.89%</td>
</tr>
<tr>
<td>ATTCGTCT</td>
<td>8.97%</td>
<td>12.79%</td>
<td>6.34%</td>
</tr>
<tr>
<td>CAAGAGA</td>
<td>7.60%</td>
<td>6.97%</td>
<td>8.18%</td>
</tr>
<tr>
<td>CACTGCC</td>
<td>11.74%</td>
<td>6.87%</td>
<td>7.06%</td>
</tr>
<tr>
<td>CCACAAA</td>
<td>7.74%</td>
<td>6.67%</td>
<td>6.16%</td>
</tr>
<tr>
<td>GCAAAGA</td>
<td>8.20%</td>
<td>7.07%</td>
<td>8.94%</td>
</tr>
<tr>
<td>GTCGTTT</td>
<td>8.97%</td>
<td>7.87%</td>
<td>8.01%</td>
</tr>
<tr>
<td>TAGCCTT</td>
<td>8.20%</td>
<td>8.53%</td>
<td>7.51%</td>
</tr>
<tr>
<td>TTCACGT</td>
<td>6.80%</td>
<td>6.67%</td>
<td>13.78%</td>
</tr>
<tr>
<td>TTCTGTG</td>
<td>8.14%</td>
<td>13.08%</td>
<td>8.04%</td>
</tr>
<tr>
<td>Illegible</td>
<td>1.05%</td>
<td>0.87%</td>
<td>0.91%</td>
</tr>
</tbody>
</table>
UltraPlex

• A unique barcoding approach that combines the strengths of past efforts without losing simplicity, efficiency, or cost effectiveness.
• Up to 9,216 samples per lane using 7bp barcodes.
• Full software support from index design through demultiplexing and reporting.
• Compatible with solution-phase as well as array-based capture.
• Offers a clinically robust indexing methodology.
UltraPlex 3D

• Extending the capabilities of the earlier Ultraplex design:
  – Up to 884,736 samples per lane
  – Routine use at 1,152 samples per lane
Simple Size Selection with beads

<table>
<thead>
<tr>
<th>Size (bp)</th>
<th>% Loss of each fragment size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ampure (ng/ul)</td>
</tr>
<tr>
<td>25</td>
<td>100%</td>
</tr>
<tr>
<td>63</td>
<td>89%</td>
</tr>
<tr>
<td>85</td>
<td>77%</td>
</tr>
<tr>
<td>100</td>
<td>83%</td>
</tr>
<tr>
<td>150</td>
<td>67%</td>
</tr>
<tr>
<td>200</td>
<td>35%</td>
</tr>
<tr>
<td>250</td>
<td>40%</td>
</tr>
<tr>
<td>300</td>
<td>68%</td>
</tr>
<tr>
<td>350</td>
<td>33%</td>
</tr>
<tr>
<td>500</td>
<td>37%</td>
</tr>
<tr>
<td>710</td>
<td>-4%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Size (bp)</th>
<th>% Loss of each fragment size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ampure (ng/ul)</td>
</tr>
<tr>
<td>25</td>
<td>100%</td>
</tr>
<tr>
<td>63</td>
<td>46%</td>
</tr>
<tr>
<td>85</td>
<td>10%</td>
</tr>
<tr>
<td>100</td>
<td>37%</td>
</tr>
<tr>
<td>150</td>
<td>-9%</td>
</tr>
<tr>
<td>200</td>
<td>-6%</td>
</tr>
<tr>
<td>250</td>
<td>-11%</td>
</tr>
<tr>
<td>300</td>
<td>-16%</td>
</tr>
<tr>
<td>350</td>
<td>-16%</td>
</tr>
<tr>
<td>500</td>
<td>-20%</td>
</tr>
<tr>
<td>710</td>
<td>-19%</td>
</tr>
</tbody>
</table>
DNA Loss over Multiple Purifications

Concentration of DNA fragments 25-766 bp before and after 5 cycles of Bead Binding/Washing/Releasing with a Single Addition of 1.8X Beads using 1.8X Binding Buffer in Each Cycle
Recent References


Acknowledgements

HudsonAlpha Institute for Biotechnology
Rick Myers
Devin Absher
Greg Barsh
Greg Cooper
Sara Cooper
Chris Gunter
Neil Lamb
Jian Han

Mayo Clinic Florida
Rosa Rademakers

Columbia University
Maria Karayiorgou
Bin Xu

University of Bergen
Stefan Johansson
Pal Njolstad

Duke University
Jennifer Perkins
Matthew Crowley

University of Michigan
Friedhelm Hildebrandt
Edgar Otto

University of Alabama-Birmingham
Lisa Guay-Woodford
Monica Tucci
Acknowledgements

**Levy Lab**
Braden Boone  
Tyson D’Angelis  
Liz Dye  
Cynthia Hendrickson  
Angela Jones  
Kruti Patel  
Terri Pointer  
Nripesh Prasad  
Melanie Robinson  
Jack Wimbish

Stephen Thomas  
Molly Delk  
Ryan Lewis  
Brooks Plummer  
Emily Fledderman  
Phil Dexheimer

**Ingenuity Systems**
Jake Leschly  
Doug Bassett  
Dan Richards  
Sandeep Sanga

**Roche NimbleGen**
Marilou Wijdicks  
Baiju Parikh  
Dan Burgess  
Dom Trillizio

**Beckman**
Alisa Jackson  
Mike Mason  
Frank Gwynn  
Andrew Kelly  
Bill Loop