Enabling True Biology with Single Molecule Sequencing

Patrice M. Milos, Ph.D.
Vice President and Chief Scientific Officer

“Sequencing, Finishing and Analysis in the Future”
DOE’s Los Alamos National Laboratory
May 27th – May 29th, 2009
A Comprehensive View of Genome Biology

- Sequencing is the method for enabling applications in:
  - Whole Genome Resequencing
  - Targeted Resequencing
  - Digital Gene Expression
  - RNA-Seq
  - Small RNA Measurements
  - Copy Number Assessment
  - Chromatin IP-Seq
  - Methylation Status

Our Understanding of Disease Requires More Than Genome Sequence
The Helicos™ Genetic Analysis System
A Production-Level Genetic Analyzer

Sample Preparation

HeliScope™ Sample Loader

HeliScope™ Single Molecule Sequencer

HeliScope™ Analysis Engine

Output

2 Flow Cells/Run
25 channels each

- Instrument ‘performance headroom’ for the $1,000 genome
- Imaging capacity ≈ 1 GB per hr
- Current chemistry > 100MB/hr
- Projected 5X chemistry improvements to 500MB/hr with existing instrument
Helicos Patented tSMS Chemistry

1. Synthesize

2. Wash

3. Image

4. Cleave

Sequencing by Synthesis
## Helicos System Performance

### Routine Usage Specifications

<table>
<thead>
<tr>
<th>Strand Output</th>
<th>12 to 16M usable strands per channel(^1) 50 Channels 600 to 800M usable strands per run</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Output</td>
<td>420 to 560 Megabases per channel 21 to 28 Gigabases per run</td>
</tr>
<tr>
<td>Throughput</td>
<td>105 to 140 Megabases per hour</td>
</tr>
<tr>
<td>Read Length</td>
<td>25 to 55 bases in length 33 to 36 average length</td>
</tr>
<tr>
<td>Accuracy</td>
<td>&gt;99.995% consensus accuracy at &gt;20X coverage</td>
</tr>
<tr>
<td>Raw Error Rate</td>
<td>≤5% (~0.2% for substitutions)</td>
</tr>
<tr>
<td>Template Size</td>
<td>25 to 5,000 bases</td>
</tr>
</tbody>
</table>

1. Usable strands are defined at ≥ 25 bases in length at the defined raw error rate
2. Dependent on applications also

---

\(^1\) For 12-mer primers,
What Differentiates True Single Molecule Sequencing (tSMS)™?

- Simplicity in Sample Prep – No PCR, No Ligations
- No Ligation, PCR for Paired Reads
- Combine Sequence and Accurate Quantitation
  - Retain Information Due to Lack of Biases
- Accuracy Throughout the Sequencing Read
- High Precision for Longitudinal Studies
  - Digital Data – Comparable Across Data Sets
- Demonstrated Sequencing of Degraded Nucleic Acid
  - FFPE DNA and RNA
  - Forensics
- Existing Methods for 1-2ng Nucleic Acid Sample Prep
  - Research Methods for 50-100 pg
# Genomic Targets – A Rapid Trajectory

<table>
<thead>
<tr>
<th>Timeframe</th>
<th>Genome</th>
<th>Size</th>
<th>Coverage</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 2007</td>
<td>M13</td>
<td>7.6kb</td>
<td>&gt;50x</td>
<td>&gt;99.5%</td>
</tr>
<tr>
<td>December 2007</td>
<td>Canine BAC</td>
<td>194kb</td>
<td>Prototype - &gt;20X</td>
<td>&gt;99.995%</td>
</tr>
<tr>
<td>May 2008</td>
<td>Yeast Transcriptome</td>
<td>&gt;6000 genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>July 2008</td>
<td><em>E. coli</em> Rhodobacter <em>Staph aureus</em></td>
<td>4.6Mb 4.3Mb 2.8Mb</td>
<td>16 Channels 48X</td>
<td></td>
</tr>
<tr>
<td>September 2008</td>
<td>Bacteria</td>
<td>“ “</td>
<td>1 Channel 80-100X</td>
<td>&gt;99.995% &gt;99.997% &gt;99.996%</td>
</tr>
<tr>
<td>October 2008</td>
<td><em>C. elegans</em></td>
<td>100Mb</td>
<td>7 Channels 27X</td>
<td>&gt;99.9995%</td>
</tr>
<tr>
<td>March 2009</td>
<td></td>
<td>3 Gb</td>
<td>3 runs 14X</td>
<td>?</td>
</tr>
</tbody>
</table>
C. elegans N2/Bristol Resequeencing Summary

- 7/50 channels loaded
- 88M reads aligned
- 2.8 GB of sequence
- 3.4% average per base error
- 0.2% sub per base
- 85% of reads 0,1,2 errors
- 27x coverage
- Variant validation
- Consensus error rate of $10^{-5}$

31M perfect reads out of 88M Aligned Reads From Seven Channels
Helicos Applications

• Bacterial Genome Sequencing: Scale and Simplicity

• Yeast Genomic Sequencing: Capturing Difficult Sequences

• Demonstrating The Power of Quantitation
  Mapping Chromatin Immunoprecipitated (ChIP) DNA
  Structural Variation: Gene Amplification
  Copy Number Variation: Origins of Replication
  Counting Human Chromosomes

• Transcriptional Profiling
  Digital Gene Expression
  RNA Seq

• Research Areas
  Small sample preparation to optimize genomics
HeliScope Workflow
Production Scale for Genomic Sequencing

- Scale from viruses to whole human genome
- Routinely able to provide 80X sequence coverage of bacterial genomes in single channels; potential for five-plex per channel with multiplex barcoding
- No bias in sequence acquisition or in quantitation due to complex preps to make the sample machine-ready
- Power to provide expression analyses at the same time
- Precision enables longitudinal studies of any application – sequence or quantitation
Even Coverage of the *E. coli* Genome

*E. coli* uniquely aligned read coverage (1 kb windows)

**Helicos**
Mean: 20.4
CV: 0.17

**Illumina**
Mean: 18.7
CV: 0.26

Identified 5 Variants from reference sequence – all five were true variants
Even Representation by Base Composition

Coverage by %GC across *E. coli* genome

Sequence coverage vs. %GC in windows

Helicos
Illumina

Aaron Berlin
How Did We Do With Other Genomes?
Similar Coverage with Differing Genomic Content

Rhodobacter Coverage

Staph Coverage
**de novo Assembly**
**Paired Read Sequencing**

- **One Approach in Product Development**
  - Library prep & ligation free *paired reads*

- **One Approach in Research feasibility studies**
  - Library prep free *paired end reads*

- **HeliScope hardware enabled for both approaches**
  - Additional reagent ports already available on instrument
    - Spacer fill nucleotides, etc.
  - Thermal control for melting & primer hybridization already available on instrument
Helicos Paired Reads – Genomic DNA
Sample Preparation – No Ligation or Amplification

Initial Studies: E coli and HapMap
Using the HeliScope Sequencer: Paired Reads
Sequence Up, Fill, Sequence Up

Step 1)
- Hybirdize DNA Template to dT<sub>50</sub> Cy5

Step 2)
- Sequence Up for 24 Quads Cy5

Step 3)
- Controlled Dark Fill Cy5

Step 4)
- Sequence Up for 24 Quads Cy5

A Unique Feature of Single Molecule Sequencing:
Useful for Small Genome Assembly,
Alternative Splicing, Translocation Identification
Genomic DNA: Paired Reads Alignments
HeliScope Sequencer - E. coli

- Initiating Genome Assembly with VELVET
- Utilizing both single and paired reads
Using Helicos Reads to Capture Unclonable Sequence

*Schizosaccharomyces octosporus* genome, 12.5 Mb
- Standard 8x Sanger draft assembly:
  570 gaps

**Approach**
- Add deep coverage of unpaired Helicos reads (assemble with Velvet)
- Attempt to close gaps with contigs
- Compare to near finished version of genome

**Results**
- Added 403,820 bases
- Closed 199 gaps (avg. 222 bp)
- Extend 174 ends (avg. 726 bp)
- Add 233kb in unanchored contigs (avg. 450 bp)

Sarah Young
Data Sets Now Available @ open.helicosbio.com

### Datasets

Listed below are a number of datasets composed of aligned reads generated with HeliScope™ technology. An overview of the data collection and analysis process is available. Data generally consist of a single channel from a HeliScope flow cell. For access to additional data from these and other experiments, send an email request. Downloads include reads in SAM and FASTA formats, along with reference sequences used for alignments.

#### Open data sets

(available to all users registered with the site)

<table>
<thead>
<tr>
<th>Species</th>
<th>Sample</th>
<th>Aligned reads</th>
<th>Download</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. elegans</em></td>
<td><em>C. elegans</em> genomic DNA</td>
<td>87,030,751</td>
<td><a href="#">C. elegans 20090116</a></td>
</tr>
<tr>
<td><em>H. sapiens</em></td>
<td>Human brain cDNA prepared from poly-A primed total RNA</td>
<td>11,663,900</td>
<td><a href="#">H. sapiens 2008_10_17_CTS4_c1_ch16_aligned.tar.gz</a></td>
</tr>
<tr>
<td></td>
<td>Human brain cDNA prepared from poly-A primed total RNA</td>
<td>12,611,637</td>
<td><a href="#">H. sapiens 2008_10_17_CTS4_c1_ch21_aligned.tar.gz</a></td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>Covari-measured E. coli genomic DNA</td>
<td>11,688,308</td>
<td><a href="#">ecoli_2008_10_08_CTS5_c12_ch3_aligned.tar.gz</a></td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>Covari-measured E. coli genomic DNA</td>
<td>11,136,748</td>
<td><a href="#">ecoli_2008_10_01_CTS4_c12_ch6_aligned.tar.gz</a></td>
</tr>
<tr>
<td><em>S. aureus</em></td>
<td>Covari-measured S. aureus genomic DNA</td>
<td>9,630,670</td>
<td><a href="#">sauereus_2008_10_01_CTS4_c12_ch11_aligned.tar.gz</a></td>
</tr>
<tr>
<td><em>R. sphaeroides</em></td>
<td>Covari-measured R. sphaeroides genomic DNA</td>
<td>7,112,158</td>
<td><a href="#">rhodo_2008_10_01_CTS4_c6_ch12_aligned.tar.gz</a></td>
</tr>
<tr>
<td><em>R. sphaeroides</em></td>
<td>Covari-measured R. sphaeroides genomic DNA</td>
<td>6,271,065</td>
<td><a href="#">rhodo_2008_10_03_CTS1_c12_ch12_aligned.tar.gz</a></td>
</tr>
</tbody>
</table>
Moving to the Human Genome
Combining Quantitation and Sequence

Providing Depth for Counting
ChIP-Seqeuencing
Collaboration with Dr. Brad Bernstein, MGH

Data Set Derived from 3-8 ng ChIP DNA

Current Method Can Utilize 250-500 pg of ChIP DNA
Assessing Copy Number Variation (CNV)
Comparison Data: Detection of Amplified Regions

- 1-2 ug DNA Sheared using Covaris
- TdT PolyA tailing
- 13 Channels HeliScope Flow Cell
- Helicos Genome Aligner
- >100M Reads Aligned

Now routinely use 50-100ng DNA
Copy Number Variation (CNV)

CNV Detection

Array CGH Data
30 Million bp

tSMS Data
2.5 Million bp

2.5 Million bp
Copy Number Variation (CNV)

Obtained ~3X Genome Coverage

Each Line is **ONE** Channel of data (3kb Smoothed)
Identifying Origins of Replication in Yeast

Hard to do:

- Can’t do comparatively: Not conserved in position or sequence
- Only been able to identify functionally
- Origins have variable efficiency

Identified in *S. pombe* and *S. cerevisiae* by cloning and selection

☞ Straightforward but laborious
☞ Method not widely applicable

☞ Can’t we just use **sequencing** as a functional assay?
Mapping DNA Replication Origins

*Schizosaccharomyces pombe* (and relatives)

1. Synchronize cells by sorting in G2
2. Grow into S phase in presence of hydroxyurea
3. Extract DNA (from G2 and S cells)
4. Sequence by Single Molecule Sequencing
5. Align reads to genome and analyze

**SMS Sequencing Allows**
- Massive number of reads at low cost
- No amplification in sample prep

Nick Rhind
Requirement to Detect Precise Genomic Content

Possible origins along genome

Actual origin usage

Peaks = X axis **position** of the origin
Height of peak on Y axis = relative **efficiency** of origin

Cell 1
Cell 2
Cell 3
Cell 4
Cell 5

Nick Rhind
Identifying Origins

Sequence alignments to *S. pombe* chromosome III

Subtract G2 from S, apply smoothing

Nick Rhind
Using Additional Functional Data to Validate Helicos data

Known origins

250bp tiling array
4 experiments

Low frequency: confirmed by DNA fiber analysis

Nick Rhind
High Resolution Read Mapping

- 250 base resolution of breakpoint of duplicated region
- Measure of accuracy of origin positions

200bp fragments used, so close to limit of resolution achievable
Can We Count Effectively at the Human Genome Level?

Experiment and Analysis Method

Sequence Two Female HapMap/ Two Male HapMap Samples – Single HeliScope Channels

- Mapped reads to human genome, discard non-unique alignments.
- Count read density in 100kb bins - Discarded bins with very low counts from all samples (repetitive sequence).
- **Normalize by sample**: Counts in each 100kb bin, genome-wide, normalize according to average bin density across all autosomes in a single sample. Measurement per chromosome is defined as the median normalized bin density across all bins in that chromosome.
- **Normalize by chromosome**: As above, then normalize values for each chromosome by the average value of that chromosome across all control samples.
Genomic DNA Samples: Assessing Sequence Read Density Across Genome Normalized By Sample, Normalized By Chromosomal Comparisons

Note: Chromosome 19 is extremely GC rich and will show altered tag density; Y chromosome tag mapping is notoriously difficult due to highly repetitive sequences.

Seq Read Density normalized by sample
Reads per 100kb per DNA sample

Seq Read Density normalized by chromosome
Reads per 100kb Normalized across samples

Increasing GC Content
Genomic DNA Analysis

- Simple, Non-PCR based SMS methods
- Sequence data shows limited bias
  - Rapid trajectory on genomic target sequencing
  - Allows equal coverage of diverse A+T, G+C rich regions
  - Coverage provides highly accurate sequence
- Attributes of SMS sequencing supports Counting
  - ChIP Seq and Copy Number Variation
  - Even distribution of sequence reads for CNV
  - Fine mapping of boundaries
  - Accurate counting of human chromosomes
  - One channel can provide high level resolution
Extending These Genomic Studies

- Optimizing SNP Sniffer Software for Variants
- Paired Reads on HapMap Samples
  - Optimize software: Both independent and dependent alignments
- Additional time course of Origins of Replication
- Continued analysis of human genome data
A View of the Transcriptome
Digital Gene Expression
Amplification-free 5’ mRNA preparation

No cDNA fragmentation
No Libraries
No PCR amplification
No PCR bias
Maintain strandedness
May allow allele specific expression
The Latest HeliScope Data

Yeast Digital Gene Expression

Channel 2 – 18.1M reads >20nt
14.6M reads >24nt

Channel 3 - 18.6M reads >20nt
14.7M reads >24nt

Demonstrated Sensitivity and Reproducibility
Allows Robust Comparisons of Transcript Differences
RNA Seq
Amplification-free 5’ mRNA Sequencing

Start with intact or fragmented RNA

cDNA Synthesis with random primers and RNA digestion

Add poly(A) tail

Hybridize & Sequence

No Libraries
No PCR amplification
No PCR bias
Maintain strandedness
May allow allele specific expression
ENCODE Program Data
RNAseq of K562 Cytosolic polyA+ RNA

181,408,862 Aligned Reads to Human Genome

- 137,317,123 Uniquely Mapping Reads
  - 19,507,924 Unique Ribosomal Reads
  - 14,052,923 Unique Mitochondrial Reads
- 103,756,276 unique reads remain
  - 83,799,334 / 103,756,276 (80.8%) reads map to exons of known genes (UCSC Known).

Describing Exonic Coverage: Total of UCSC 238,209 projected exons

- 78,602 exons are 100% covered by our reads
- 108,458 exons are at least 90% covered
- 117,403 exons are at least 80% covered
- 133,824 exons are at least 50% covered
- 74,702 have no coverage
Characterized human exons of SLC25A1 gene on Chromosome 22
Novel transcription units outside of the SLC25A1/SLC25A1 intron
Differences Between Individuals: TRIM14 locus.
Prototype Fusion Detection Algorithm: Overview

1. Align w/ SW

2. Cluster by breakpoint; check overhang consistency

min 18bp

clusters

3. Match left & right breakpoints

4. Realign unaligneds to fusion sequence

Align to TXome, genome

unaligned
Example

- **BCR-ABL fusion transcript in K562 cell-line**
- **Breakpoint is covered by ~50 reads**

**Key:**
- **Left match**
- **Right match**
- **Post match**
Novel Fusion Event Identified

K562, 5 channels

CAACCTCTGGGTTTCAGCTTTTGCCAAGCTTTCAGCACC-TGTAAG consensus

CAACCTCTGGGTTTCAGCTTTTGCCAAGCTTTCAGCACC-TGTAAG Novel Gene Fusion Partner
Transcriptome: Paired Read Alignments
HeliScope Sequencer – K562 Transcriptome

Paired Sequence Reads
Identification and Characterization of Transcript Variants
RNA Methods
Digital Gene Expression and Whole Transcriptome Resequencing

- Simple, Non-PCR based SMS methods
- Digital Gene Expression
  - Provides single sequence tag per transcript
  - Provided details on transcription start sites
- Whole Transcriptome Resequencing
  - Maintains strandedness
  - Variety of methods for desired results
  - Method for even distribution of sequence reads
  - Useful for transcript splicing and fusion gene identification
  - Low false positive results in fusion studies
- Small RNA Measurements
  - Method finalized for unamplified small RNA Seq
Helicos Research Methods
Optimizing Methods for Minimal Sample Use

- **Small Sample Preparation**
  - DNA sequencing: 150 pg tailed and sequenced
  - RNA Seq: 500 pg/7M aligned reads
  - Digital Gene Expression: 2 ng HeLa total RNA/12 M aligned reads (40% ribo/mito RNA)

- **FFPE DNA Sequencing**: Utilizing 5-10 ng
- **FFPE RNA Seq**: Utilizing 100ng total RNA
- **Direct RNA Sequencing**: 2 pg
Sequencing is the method for enabling applications in:

- Whole Genome Resequencing
- Targeted Resequencing
- Digital Gene Expression
- RNA-Sequencing
- Small RNA Measurements
- Copy Number Assessment
- Chromatin IP-Sequencing
- Methylation Status

Our Understanding of Disease Requires More Than Genome Sequence
Acknowledgments

Broad
Chad Nusbaum
Carsten Russ
Aaron Berlin
Sara Young
Numerous Colleagues

U Mass Worcester
Nick Rhind and
Colleagues

MGH
Brad Bernstein

Boston College
Gabor Marth
Chip Stewart

NYU
David Fitch
Karin Kiontke

CSHL
Tom Gingeras

Helicos Colleagues

Special Thanks For Funding
NHGRI

NHGRI
Mike Erdos
Francis Collins
Variant Detection: Mutation Finding
Collaboration with Albert Einstein College of Medicine

Genomic Samples: Breast Cancer Cell Lines

Genes Of Interest

- BRCA1, BRCA2, ATM, CHK1, CHK2, FGFR2, p53

Long Range PCR Products Provided to Helicos for Targeted Resequencing

- Sequence each sample using single channel on HeliScope
- Align to whole human genome to define gene boundaries
- Realign to genome regions
- Utilize Helicos SNP Finding Tool for SNP and Mutation Detection
Helicos Software For Variant Detection
snpSniffer structure

- Determine error rate
- Alignments
- CREATE COVERAGE SUMMARY
- ANLAYS EACH ROW IN COVERAGE SUMMARY FOR PRESENCE OF SNPS
- RE-ALIGNMENT/RE-ANALYSIS MODULE
- CONFIRMED SNP ?
- SNP REPORT

Based on forward and reverse alignments
Exonic Coverage for Mutation Detection

- All seven genes were successfully sequenced to ~50-100X coverage, more than sufficient for variant detection.
  - Majority of exons in excess of 100x coverage.

- SNP detection provided list of variants in each sample.
  - snpSniffer optimized for substitution SNPs where it appears to have provided hundreds of high confident SNP calls.
PCR and Sequencing Reactions Behaved Similarly Across Samples
### Excerpt of the SNP Table

#### High Confidence Mutation Discovery

<table>
<thead>
<tr>
<th>Ref Name</th>
<th>Chromosome</th>
<th>Position</th>
<th>Type</th>
<th>Change</th>
<th>P-value</th>
<th>A</th>
<th>C</th>
<th>T</th>
<th>G</th>
<th>G - knownSNP</th>
<th>Left Flanking</th>
<th>SNP</th>
<th>Right Flanking</th>
</tr>
</thead>
<tbody>
<tr>
<td>p53</td>
<td>chr17</td>
<td>7517846</td>
<td>G-&gt;A</td>
<td>4.83E-290</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHK1</td>
<td>chr11</td>
<td>125018307</td>
<td>G-&gt;A</td>
<td>5.25E-288</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>rsID:79519</td>
<td>TGGCAATAGGAGCTGGCTTGGCCGCCG</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>BRAC2</td>
<td>chr13</td>
<td>31813005</td>
<td>G-&gt;C</td>
<td>4.36E-286</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>rsID:20607</td>
<td>TACATTGGTCTGAGGTTTGAATCTTGGT</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>ATM</td>
<td>chr11</td>
<td>107688377</td>
<td>A-&gt;G</td>
<td>3.63E-285</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>rsID:65924</td>
<td>TCTGACATTGGAGCTGGCTTGGCCGCCG</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>ATM</td>
<td>chr11</td>
<td>107710593</td>
<td>A-&gt;A</td>
<td>3.12E-284</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>rsID:22706</td>
<td>CATACATTGGAGCTGGCTTGGCCGCCG</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>ATM</td>
<td>chr11</td>
<td>107648392</td>
<td>C&gt;T</td>
<td>1.59E-282</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>rsID:66467</td>
<td>AGAAAGACATTGGAGCTGGCTTGGCCGCCG</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>ATM</td>
<td>chr11</td>
<td>107699283</td>
<td>C&gt;T</td>
<td>1.59E-282</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>rsID:59574</td>
<td>AGAAAGACATTGGAGCTGGCTTGGCCGCCG</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>FGFR2</td>
<td>chr10</td>
<td>123347551</td>
<td>C&gt;T</td>
<td>1.59E-282</td>
<td>0.01</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>rsID:10471</td>
<td>GAGGAGGAGACATGGAGCTGGCTTGGCCGCCG</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>CHK1</td>
<td>chr11</td>
<td>125001666</td>
<td>G-&gt;T</td>
<td>2.54E-281</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>rsID:62386</td>
<td>ATGTTAATTTTCCATGAGGTTTGAATCTTGGT</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>ATM</td>
<td>chr11</td>
<td>107648211</td>
<td>C&gt;T</td>
<td>1.17E-281</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>rsID:63706</td>
<td>ATACATTAGTCGTGAGGTTTGAATCTTGGT</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>ATM</td>
<td>chr11</td>
<td>107611992</td>
<td>C&gt;T</td>
<td>2.33E-281</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>rsID:62386</td>
<td>ATACATTAGTCGTGAGGTTTGAATCTTGGT</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>p53</td>
<td>chr17</td>
<td>7517747</td>
<td>G-&gt;A</td>
<td>3.91E-283</td>
<td>0.99</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>rsID:20609</td>
<td>CCTCCAGTCAGCTGAGGTTTGAATCTTGGT</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>ATC1</td>
<td>chr17</td>
<td>38469351</td>
<td>C&gt;T</td>
<td>9.11E-282</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>rsID:66467</td>
<td>AGAAAGACATTGGAGCTGGCTTGGCCGCCG</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>ATM</td>
<td>chr11</td>
<td>107648392</td>
<td>C&gt;T</td>
<td>2.54E-282</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>rsID:65924</td>
<td>AGAAAGACATTGGAGCTGGCTTGGCCGCCG</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>ATM</td>
<td>chr13</td>
<td>31843932</td>
<td>C&gt;T</td>
<td>5.78E-277</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>rsID:57301</td>
<td>AGAAAGACATTGGAGCTGGCTTGGCCGCCG</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>ATM</td>
<td>chr11</td>
<td>107744638</td>
<td>C&gt;T</td>
<td>1.75E-287</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>rsID:4585</td>
<td>AGAAAGACATTGGAGCTGGCTTGGCCGCCG</td>
<td>A</td>
<td></td>
</tr>
</tbody>
</table>