Detecting Cancer-Associated Structural Variants Using Megabase-Scale DNA

James R. Broach
Penn State College of Medicine

Bionano Webinar
May 11, 2017
Molecular Classification of AML and Overall Survival

Papaemmanuil et al. (2016) NEJM 374: 2209
The Next, Next Generation: Building an Individualized Genome

The Goal:
Build individualized cancer genome assemblies *de novo* (i.e. without the use of a reference genome)

The Challenge:
Repetitive sequences limits the ability to assemble cancer genomes based on next-gen sequence data alone
The Bionano System

Instrument

Consumable

Reagents

Saphyr Chip™

Bionano Prep™ kit

IrysChip®

Software

Bionano Access™

Irys®
Bionano Chip – NanoChannel arrays on silicon

Leverages mature semiconductor manufacturing
- High-quality wafer-scale manufacturing in state-of-the-art semiconductor facility

Thousands of parallel NanoChannels arrays
- Imparts uniform and orderly format for accurate measurements

Chip schematic

- High-throughput cartridge
- Multiple

DNA flow

Saphyr Chip™

IrysChip®
De Novo Cancer Genome Assembly
Mapping the Cancer Genome
Mapping the Cancer Genome
Mapping the Cancer Genome

**SV identification**
- Hi-C
- Irys
- WGS

**Validation**
- FISH
- RT Timing
- PCR-Sanger

**Comparative analysis**
- Technique-specific advantages
- Predicting unresolved genome regions

**Functional analysis**
- Coding region disruption:
  - Fusion transcripts
  - Copy number variation
  - Gene truncation
- Regulatory element disruption:
  - SVs disrupt promoters
  - SVs disrupt enhancers
  - SVs alter TAD domain
Chromosome Conformation Capture For Translocation Detection

Usually within the same chromosomes

Two ends can come from two separate regions of one chromosome
Mapping the Cancer Genome

Caki2: TL detected by Hi-C

TL detected by Irys

Cancer genome contig

Single DNA molecule

TL detected by WGS

WGS reads

chr 3

chr 2

chr 3

chr 2
Three Methods Identified Novel Translocations Additional To Known Ones

T47D:
Novel and known inter-chr TLs

<table>
<thead>
<tr>
<th>Translocations</th>
<th>Previous karyotype</th>
<th>Detected by our methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr6 chr22</td>
<td>Novel</td>
<td>Hi-C, WGS</td>
</tr>
<tr>
<td>chr4 chr5</td>
<td>Novel</td>
<td>Hi-C, Irys, WGS</td>
</tr>
<tr>
<td>chr3 chr9</td>
<td>Novel</td>
<td>Hi-C, Irys, WGS</td>
</tr>
<tr>
<td>chr3 chr10</td>
<td>Novel</td>
<td>Hi-C, WGS</td>
</tr>
<tr>
<td>chr3 chr10</td>
<td>Novel</td>
<td>Hi-C, WGS</td>
</tr>
<tr>
<td>chr12 chr20</td>
<td>Novel</td>
<td>Hi-C, Irys, WGS</td>
</tr>
<tr>
<td>chr9 chr15</td>
<td>Novel</td>
<td>Hi-C, Irys, WGS</td>
</tr>
<tr>
<td>+ 52 other translocations + 13 intra-chromosomal translocations</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

chr3 chr5 known Hi-C, Irys
chr3 chr10 known Hi-C, Irys
chr3 chr12 known Hi-C, Irys, WGS
chr6 chrX known Hi-C
chr7 chr15 known Hi-C, WGS
chr8 chr14 known Hi-C, Irys
chr9 chr17 known Hi-C
chr10 chr20 known Hi-C, Irys, WGS
chr10 chr20 known Irys
chr12 chr13 known Hi-C
chr12 chr16 known Hi-C
chr11 chr17 Known -
chr1 chr16 known -
Optical mapping has higher sensitivity to deletions enriched of repetitive elements.
Optical mapping detects larger SVs

T47D deletion

<table>
<thead>
<tr>
<th>Method</th>
<th>Count</th>
<th>(Size)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irys</td>
<td>206</td>
<td></td>
</tr>
<tr>
<td>WGS</td>
<td></td>
<td>(298)</td>
</tr>
</tbody>
</table>

SV size (log10)

- Irys Only
- Overlap
- WGS only

Graph showing distribution of SV sizes.
Cancer cell exhibits more rare and novel deletion, enriched of cancer-related genes

Optical mapping

WGS

Intersect

Compare to previously reported polymorphic SVs

Previously reported polymorphism

Unreported

Reported frequency

T47D
K562
Caki2
GM12878

ACSL3
ERC1
MSI2
HLA-A
LHFP
CASP8
MAP3K13
GPHN
IKKKB
LPP, PDE4DIP
BCL9
KEAP1, LPP
ERC1
TBL1XR1
RNF213
FANCC
Major contribution by larger deletions to cell-type specific cancer

**Optical mapping** → **WGS** → **Intersect** → **Functional enrichment analysis**

### T47D genome-wide deletion
- DNA recombination
- Genes down-regulated in MCF7 cells
- Human breast tumor proliferation
- Genes upregulated in primary breast cancer overexpressing E2F3
- Positive regulation of mammary gland epithelial cell proliferation
- Genes upregulated during pubertal mammary gland development
- Genes up/down-regulated in basal/luminal mammary epithelium

### Caki2 genome-wide deletion
- Glutathione binding
- Glutathione transferase activity
- Cellular response to mineralocorticoid stimulus
- Double-stranded telomeric DNA binding
- Renal-urinary system mesenchyme
- Regulation of lipid transport
- Neutral amino acid transmembrane transporter activity
- Genes downregulated in kidney fibroblasts expressing active CTNNB1

### K562 genome-wide deletion
- Genes upregulated in T lymphocyte
- Double-stranded telomeric DNA binding
- Hematopoietic stem cell differentiation
- Genes upregulated in neutrophils
- B Lymphocyte Cell Surface Molecules
- Genes upregulated in B-lymphoma
- IL 4 signaling pathway
- Increased erythroid progenitor cell number
- Defective B cell activation
- p53 signaling pathway
Building Cancer Genome Profiles

Copy number alteration, translocation, and chromothripsis are hallmarks of cancer cell lines

Kidney cancer: Caki2    Chronic Leukemia: K56    Breast cancer cell line: T47D

Plotted translocations are detected by 2 methods or more.

Normal cell line: NA12878
De Novo Assembly of a Leukemia Genome

Test Case:
AML (Patient Derived)
46XY, t(6;9), t(8;13)

Question:
Can we identify known translocations using single-molecule optical mapping?

De Novo Assembly of a Leukemia Genome

De Novo Assembly: ~100X Coverage
46XY

Chr8 Chr13

PennState
The Milton S. Hershey Medical Center
The College of Medicine
De Novo Assembly of a Leukemia Genome
2yo F with T-cell ALL 46,XX,add(9)(p13)[7]/46,XX[13] STIL gene deletion with retention of 1780 Insertions

TAL1 (at 1p32) 823 Deletions
CDKN2A deletion 9(p21) 46 Inversions
1 Translocation
Cytogenetic Karyotype:
46,XX,add(9)(p13)[7]/46,XX[13]
Additional data:
CDKN2A Deletion
Bionano Karyotype:
Effective coverage 90x
XX,t(1;6),t(9;16),del9(CDKN2A)
Cytogenetic Karyotype:
46,XY,del(6)(q13q23),add(12)
(p11.2)[17]/46,XY[3]
Additional data:
CDKN2A deletion; ETV6-RUNX1-
positive [fusion] t(12;21)

Bionano Karyotype:
Effective coverage 83-101x
XY,t(1;6),t(1;9),t(8;17),t(9;16),
t(12;21),add(8),del(6),
del(CDKN2A)
Cytogenetic Karyotype:
47,XX,+8[16]/46,XX[4]

Bionano Karyotype:
Effective coverage 106x
XX,t(1,9),add(8)
Cytogenetic Karyotype:
46,XY,t(4;19)(q12;q13.3)?c[20]

Bionano Karyotype:
Effective coverage 98x
XY,t(1;9),t(1;20),t(4;19),del(4)
Leukemia 1021

Cytogenetic Karyotype:
46,XY,t(6;9)(p23;q34),t(8;13;11)
(q22;q12;q23)[20]
## De Novo Assembly of Leukemia Genome Summary to Date

<table>
<thead>
<tr>
<th>Patient</th>
<th>Cytogenetics</th>
<th>Additional data</th>
<th>Bionano</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 yo F T-cell ALL</td>
<td>46,XX,add(9)(p13)[7]/46,XX[13]</td>
<td>STIL deletion; CNKN2A deletion</td>
<td>t(1:16), 46 inversions, 823 deletions</td>
</tr>
<tr>
<td>10 yo F AML</td>
<td>47,XX,+8[16]/46,XX[4]</td>
<td>Trisomy 8</td>
<td>t(8:20),t(2:8),t(5:15), t(7:10),t(9:12),t(13:17)</td>
</tr>
<tr>
<td>55 yo M AML</td>
<td>46,XY,t(8;21)(q22;q22)[20]</td>
<td>NA</td>
<td>t(8:21), 92 inversions, 1636 deletions</td>
</tr>
<tr>
<td>63 yo F AML</td>
<td>46,XX, der(7)t(7;11)(q35;q12),inv(16)</td>
<td>MYH11/CBFB fusion FLT3 TKD Mutation</td>
<td>t(7:11),t(16:16), 72 inversions, 1655 dels</td>
</tr>
<tr>
<td>29 yo F AML</td>
<td>46,XY,t(9;22)(q34;q11.2)[20]</td>
<td>Elevated BCR-ABL p210 mRNA</td>
<td>t(9:22), 60 inversions, 1517 deletions</td>
</tr>
</tbody>
</table>
# De Novo Assembly of Leukemia Genome Summary to Date

<table>
<thead>
<tr>
<th>Patient</th>
<th>Cytogenetics</th>
<th>Additional data</th>
<th>Bionano</th>
</tr>
</thead>
<tbody>
<tr>
<td>71 yo F B-cell ALL</td>
<td>49,X,der(X)t(X;1)(q13;q11),del(3)(q21),add(4)(q31.3),add(6)(q23),-8,add(9)(p11),add(12)(q22),del(13)(q12q14),add(14)(q32),add(15)(p11.1),-18,+5mar[11]/46,XX[9]</td>
<td>MYC rearrangement BCL2/IGH fusion</td>
<td>t(3:6),t(3:6),t(3:4),t(3:8),t(4:8),t(4:18),t(6:8),t(8:9),t(8:18),t(9:18),t(15:18),74 inversions, 1682 deletions</td>
</tr>
<tr>
<td>55 yo M AML</td>
<td>44,XY,del(1)(q42),-7,add(8)(p23),-12[18]/44,idem,t(4;5)(p14;q13)[2]</td>
<td></td>
<td>t(1:12),t(8:12),80 inversions,1574 deletions</td>
</tr>
<tr>
<td>58 yo F AML</td>
<td>46,XY,t(6;9)(p23;q34),t(8;13;11)(q22;q12;q23)[20]</td>
<td>DEK/NUP214 fusion</td>
<td>t(8:13),t(6:9),t(9:6)</td>
</tr>
</tbody>
</table>
Conclusions & Implications

• “Next Generation” Genetics & Genomics
  – Identify novel cancer predisposing mutations
  – Generate some of the first patient-derived cancer genome assemblies

• Combining current sequencing platforms with novel genomics techniques will allow unprecedented insight into cancer genomics

• High-throughput genome mapping could potentially replace karyotyping for translocation identification
Acknowledgements

Penn State Institute for Personalized Medicine
  Feng Yue
  Jie Yu
  Yanli Wang
  Fan Song

Penn State Department of Otolaryngology
  Christopher Pool
  Darrin Bann

Penn State Department of Hematological Oncology
  David Claxton
  Shin Mineishi

Bionano Genomics
  Alex Hastie
  Benjamin Clifford
  Weiping Wang
  Ahmed Naguib

U Mass Med School
  Job Dekker
De Novo Assembly of Leukemia Genome

peds1907

peds1

peds1916
De Novo Assembly of Leukemia Genome

NA12878
peds1953
Leukemia1021
De Novo Assembly of Leukemia Genome

BN936
BN1160
BN784