The Next Generation of Gene Fusion Discovery in Cancer

Arul M. Chinnaiyan, M.D., Ph.D.
Departments of Pathology and Urology
Howard Hughes Medical Institute
Areas where the Cancer DBP will interact with the NCIBI

- Bioinformatics related to metabolomics
- Integrative analysis across molecular alterations
- Bioinformatics related to Next Generation sequencing
Metabolomic Profiling of Cancer Progression

Complexity of the -omics

Methylation

- DNA
- RNA
- Protein
- Biochemicals (Metabolites)

Epigenetics

- Genomics – 25,000 Genes
- Transcriptomics – 100,000 Transcripts
- Proteomics – 1,000,000 Proteins
- Metabolomics – 2,400 Compounds
Genomic Loss of microRNA-101 Leads to Overexpression of Histone Methyltransferase EZH2 in Cancer

Sooryanarayana Varambally, Qi Cao, Ram-Shankar Mani, Sunita Shankar, Xiaosong Wang, Bushra Ateeq, Bharathi Laxman, Xuhong Cao, Xiaojun Jing, Kalpana Ramnarayanan, J. Chad Brenner, Jindan Yu, Jung H. Kim, Bo Han, Patrick Tan, Chandan Kumar-Sinha, Robert J. Lonigro, Nallasivam Palanisamy, Christopher A. Maher, Arul M. Chinnaiyan

Nov 2008
The integrative model for translation of bio-data into novel gene fusions
Breaking the rules of cancer

David R Shaffer & Pier Paolo Pandolfi

A cancer genetics dogma states that hematologic malignancies arise as a result of defined chromosomal translocations, whereas mutations underlie epithelial solid tumors. This rule is now broken in an analysis of chromosomal translocations in prostate cancer.
BCR-ABL Gene Fusion in CML (a type of leukemia)

http://www.cmlsupport.com/bcrabl.gif
A Computational Approach Leads to the Discovery of Gene Fusions in Prostate Cancer

TMPRSS2-ETS Gene Fusions in Prostate Cancer

Tomlins et al. Science 310:644
Confirmation of High Prevalence of Gene Fusions (40-80%)
Appearance of Gene Fusions in Prostate Cancer Progression

- Gene fusions occur in HG-PIN contiguous or in close proximity to prostate cancer
- Multi-focal nature of prostate cancer
- Clonal nature of gene fusions
ETS Gene Fusion Products Induce Cell Invasion

[Graphs showing comparison of RWPE Stable Lines and Primary Prostate Epithelial Cells (Transient transfection)]

Control  ERG  ETV1

Control  ERG  ETV1
Knockdown of TMPRSS2-ERG gene fusion inhibits cell invasion, proliferation and tumor growth

VCaP cells (TMPRSS2-ERG +)

VCaP Xenografts

Shiv Srivastava et al
Oncogene (2008) 27:5348
Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate

Brett S Carver1,2, Jennifer Tran1, Anuradha Gopalan3, Zhenbang Chen1,4, Safa Shaikh2, Arkaitz Carracedo1,4, Andrea Alimonti1,4, Caterina Nardella1,4, Shohreh Varmeh1,4, Peter T Scardino2, Carlos Cordon-Cardo5, William Gerald5 & Pier Paolo Pandolfi1,3,4
A Family of Gene Fusions
Molecular Sub-Types
in Prostate Cancer

Kumar-Sinha et al.
Nature Cancer Reviews
2008
Molecular Subtypes of Prostate Cancer

• ~70% of North American prostate cancers have ETS gene fusions
  50-60% TMPRSS2-ERG
  TMPRSS2-ERG with deletion (50-60%)
  TMPRSS2-ERG without deletion (40-50%)
  About 15 variant fusion transcripts
  ~5-10% ETV1 Fusions
  ~1% ETV4 Fusions
  ~1% ETV5 Fusions

• ~30% Negative for ETS gene fusions
Clinical implications for prostate cancer diagnosis?
Gene fusion urine test

Specimen collection and assay format

Quantitative measurement of TMPRSS2-ERG mRNA in post-DRE sediments and whole urine
TMPRSS2:ERG in the urine of men with prostate cancer

Similar results (cancer vs. benign, significant vs. insignificant) in SDVA/UL cohort
Prostate cancer therapy: How do we target the prostate cancer gene fusions?
A Molecular Basis for Prostate Cancer

AR = androgen receptor
ARE = androgen response element
DHT = dihydrotestosterone
ETS = ETS family of transcription factors (ERG/ETV1)

Diagram:
- Anti-Androgens
- DHT
- AR
- TMPRSS2 (ARE)
- ETS Genes
- ETS Target Genes:
  - Cell growth
  - Invasion/Metastasis
  - Cell Survival
Subversion of Tissue-Specific Promoters/Enhancers to Cause Cancer

- IgH
- BCL2
- TCR
- MYC
- TMPRSS2 (ARE)
- ETS Genes
- ERE
- ???
- ???
- ???
- B cell lymphomas
- T cell lymphomas
- Prostate Cancer
- Breast Cancer?
- Gynecologic Cancer?
- Others?
  (e.g., Lung, Colon, Ovarian, Liver)
Identification of the transforming EML4–ALK fusion gene in non-small-cell lung cancer

Manabu Soda¹,², Young Lim Choi¹, Munehiro Enomoto¹,², Shuji Takada¹, Yoshihiro Yamashita¹, Shunpei Ishikawa⁵, Shin-ichiro Fujiwara¹, Hideki Watanabe¹, Kentaro Kurashina¹, Hisashi Hatanaka¹, Masashi Bando², Shoji Ohno³, Yuichi Ishikawa⁶, Hiroyuki Aburatani⁵,⁷, Toshiro Niki³, Yasunori Sohara⁴, Yukihiko Sugiyama² & Hiroyuki Mano¹,⁷

¹Division of Functional Genomics, ²Division of Pulmonary Medicine, ³Department of Pathology, and ⁴Division of General Thoracic Surgery, Jichi Medical University, Tochigi 329-0048, Japan. ⁵Research Center for Advanced Science and Technology, University of Tokyo, Tokyo 153-8904, Japan. ⁶Department of Pathology, The Cancer Institute, Japanese Foundation for Cancer Research, Tokyo 135-8550, Japan. ⁷Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency, Saitama 332-0012, Japan.
Conclusions-1

- **Majority** of prostate cancers have a gene fusion/translocation
- ETS rearrangements in prostate cancer are the **most common gene fusion** in cancer
- We propose that gene fusions of *prostate-active genomic regulatory regions* to *oncogenic factors* is the basis for the development of prostate cancer
- Gene fusions may serve as useful **biomarkers** of prostate cancer (diagnosis and prognosis)
- Gene fusions may represent a **rational therapeutic target** for prostate cancer (in vitro and in vivo models of gene fusion)
- **Family of 5’ fusion partners** identified with functionally different upstream regulatory elements
- Suggest that **other common solid tumors** may be the result of thus far unidentified recurrent gene fusions (buried in the noise of non-specific alterations)
Nextgen Transcriptome Sequencing to Detect Gene Fusions in Cancer
Next Gen sequencing

Applied Biosystems
ABI 3730XL
1 Mb / day

Roche / 454
Genome Sequencer FLX
100 Mb / run

Illumina / Solexa
Genetic Analyzer
2000 Mb / run

Reference Genome
Massively Parallel, High throughput, Next-Gen sequencing
Next Generation Sequencing Fusion Pipeline

COPA; Oncomine

Exon-walking (qPCR)

Fusion-specific PCR

FISH

(Kumar-Sinha, et al., Nature Cancer Reviews 2008)
Next Generation Transcriptome Sequencing for gene fusion discovery

- Circumvents limitations introduced by cytogenetic techniques
- Increased sensitivity over microarrays
- Unbiased view of expressed portion of the genome
- Requires less resources than genomic sequencing, while focusing on aberrations are primarily expected to have functional consequences in carcinogenesis
- Provides evidence of both partners within fusion event
- Gene fusions as a class of mutations have been difficult to study comprehensively relation to DNA substitution, amplification and deletion
Chimera discovery pipeline

Sequence long reads (200-500 nt)

Category 1: Mapping reads

Category 2: Partially aligned reads

Category 3: Non-mapping reads

Viral
Contamination
Other

Chimera between distant genes

Chimera between neighboring genes

Inter-chromosomal chimeras

Intra-chromosomal chimeras

Gene fusion candidate

Gene fusion candidate

Read-through candidate

Gene x
Gene y
Gene z

Gene x
Gene y
Gene z

Maher et al, Nature 2009
“Re-discovery” of BCR-ABL1

- Transcriptome sequencing of chronic myelogenous leukemia cell line (K562)
- ~70 million 36-mer sequence reads
- 19 chimeras spanning BCR-ABL1 fusion boundary

Genbank accession: M30829

Chromosome 22

Chromosome 9

BCR

ABL1

Chromosome 22

Chromosome 9
“Re-discovery” of TMPRSS2-ERG

- ~76 million 36-mer sequence reads from 8 lanes
- 21 chimeras spanning TMPRSS2-ERG fusion boundary in VCaP cells
<table>
<thead>
<tr>
<th>Chimera</th>
<th>Chimera Class</th>
<th>Location</th>
<th>3' Gene</th>
<th>Location</th>
<th>3' Gene</th>
<th>Validated in</th>
<th>Validated by</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCR-ABL1</td>
<td>Class I: Translocation</td>
<td>22q11.23</td>
<td>BCR, breakpoint cluster region</td>
<td>9q34.1</td>
<td>ABL1, c-abl oncogene 1, receptor</td>
<td>K562</td>
<td>Short read, qRT-PCR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>tyrosine kinase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRPS10-HPR</td>
<td>Class I: Translocation</td>
<td>6p21.1</td>
<td>MRPS10, mitochondrial ribosomal protein S10</td>
<td>16q22.1</td>
<td>HPR, g-protein-related protein</td>
<td>LNCaP</td>
<td>Long read, Short read, qRT-PCR, FISH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIPOL1-DGKB</td>
<td>Class II: Interchromosomal</td>
<td>14q13.3-q21.1</td>
<td>MIPOL1, mirror-image polydactyly 1</td>
<td>7p21.2</td>
<td>DGKB, diacylglycerol kinase, beta</td>
<td>LNCaP</td>
<td>Long read, Short read, qRT-PCR, FISH</td>
</tr>
<tr>
<td></td>
<td>complex</td>
<td></td>
<td></td>
<td></td>
<td>90kDa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMPRSS2-ERG*</td>
<td>Class III: Intestinal</td>
<td>21q22.3</td>
<td>TMPRSS2, transmembrane protease, serine 2</td>
<td>21q22.3</td>
<td>ERG, v-ets erythroblastosis virus E26</td>
<td>VCaP, VCaP-Met</td>
<td>Long read, Short read, qRT-PCR, FISH</td>
</tr>
<tr>
<td></td>
<td>Deletion</td>
<td></td>
<td></td>
<td></td>
<td>oncogene homolog (avian)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP10-ZDHHC7*</td>
<td>Class III: Intestinal</td>
<td>16q24.1</td>
<td>USP10, ubiquitin specific peptidase 10</td>
<td>16q24.1</td>
<td>ZDHHC7, zinc finger, DHHC-type</td>
<td>VCaP, VCaP-Met</td>
<td>Long read, Short read, qRT-PCR, FISH</td>
</tr>
<tr>
<td></td>
<td>Deletion</td>
<td></td>
<td></td>
<td></td>
<td>containing 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STRN4-GPSN2*</td>
<td>Class IV: Intrachromosomal</td>
<td>19q13.2</td>
<td>STRN4, striatin, calmodulin binding protein 4</td>
<td>19p13.12</td>
<td>GPSN2, glycoprotein, synaptic 2</td>
<td>Met-3</td>
<td>Short read, qRT-PCR</td>
</tr>
<tr>
<td></td>
<td>complex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LMAN2-AP3S1</td>
<td>Class IV: Intrachromosomal</td>
<td>5q35.3</td>
<td>LMAN2 lectin, mannoside-binding 2</td>
<td>5q22</td>
<td>AP3S1, adaptor-related protein complex</td>
<td>VCaP, VCaP-Met</td>
<td>Short read, qRT-PCR</td>
</tr>
<tr>
<td></td>
<td>complex</td>
<td></td>
<td></td>
<td></td>
<td>3, sigma 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HJURP-EIF4E2*</td>
<td>Class IV: Intrachromosomal</td>
<td>2q37.1</td>
<td>HJURP, Holliday junction recognition protein</td>
<td>2q37.1</td>
<td>EIF4E2, eukaryotic translation initiation factor 4E family member 2</td>
<td>VCaP, VCaP-Met</td>
<td>Long read, Short read, qRT-PCR, FISH</td>
</tr>
<tr>
<td></td>
<td>complex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INPP4A-HJURP*</td>
<td>Class II: Intrachromosomal</td>
<td>2q11.2</td>
<td>INPP4A, inositol polyphosphate-4-phosphatase, type I</td>
<td>2q37.1</td>
<td>HJURP, Holliday junction recognition</td>
<td>VCaP, VCaP-Met</td>
<td>Long read, Short read, qRT-PCR, FISH</td>
</tr>
<tr>
<td></td>
<td>complex</td>
<td></td>
<td></td>
<td></td>
<td>protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC3H2-RGS3</td>
<td>Class IV: Intrachromosomal</td>
<td>6q34</td>
<td>RC3H2, ring finger and CCCH-type zinc finger domains 2</td>
<td>9q32</td>
<td>RGS3, regulator of G-protein signaling</td>
<td>VCaP, VCaP-Met</td>
<td>Short read, qRT-PCR</td>
</tr>
<tr>
<td></td>
<td>complex</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZNF649-ZNF577</td>
<td>Class V: Read-through</td>
<td>19q13.33</td>
<td>ZNF649, zinc finger protein 649</td>
<td>19q13.33</td>
<td>ZNF577, zinc finger protein 577</td>
<td>VCaP, VCaP-Met</td>
<td>Long read, Short read, qRT-PCR</td>
</tr>
<tr>
<td>MBTPS2-YY2*</td>
<td>Class V Read-through</td>
<td>Xp22.1-p22.2</td>
<td>MBTPS2, membrane-bound transcription factor peptidase, site 2</td>
<td>Xp22.2-p22.1</td>
<td>YY2, YY2 transcription factor</td>
<td>VCaP, LNCaP, VCaP-Met</td>
<td>Long read, Short read, qRT-PCR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C19orf25-APC2</td>
<td>Class V: Read-through</td>
<td>19p13.3</td>
<td>C19orf25, chromosome 19 open reading frame 25</td>
<td>19p13.3</td>
<td>APC2, adenomatous polyposis coli 2</td>
<td>LNCaP</td>
<td>Long read, Short read, qRT-PCR</td>
</tr>
<tr>
<td>WDR55-DND1</td>
<td>Class V: Read-through</td>
<td>5q31.3</td>
<td>WDR55, WD repeat domain 55</td>
<td>5q31.3</td>
<td>DND1, dead end homolog 1 (zebrafish)</td>
<td>RWPE</td>
<td>Long read, Short read, qRT-PCR</td>
</tr>
<tr>
<td>SLC45A3-ELK4*</td>
<td>Class V: Read-through</td>
<td>1q32.1</td>
<td>SLC45A3, solute carrier family 45 member 3</td>
<td>1q32.1</td>
<td>ELK4, ETS domain-containing protein</td>
<td>Mel-4</td>
<td>Short read, qRT-PCR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Read-through

The product of co-transcription of adjacent genes coupled with intergenic splicing (CoTIS)

<table>
<thead>
<tr>
<th>Gene fusions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>VCaP</td>
<td>1</td>
</tr>
<tr>
<td>LNCaP</td>
<td>3</td>
</tr>
<tr>
<td>RWPE</td>
<td>1</td>
</tr>
</tbody>
</table>
Read-through Chimeric transcripts

VCaP

LNCaP

RWPE

Chr 19

ZNF577

ZNF649

APC2

C19orf25

Chr 5

WDR55

DND1
SLC45A3-ELK4: a Novel, Recurrent, Prostate Specific, Androgen sensitive Chimera

solute carrier family 45, member 3 (Prostate cancer associated protein, Prostein)

ELK4, ETS-domain protein
(SRF accessory protein 1)

**LNCaP, SLC45A3-ELK4**

Target/GAPDH

- Urx
- R1881 24h
- R1881 48h
Lack of microdeletions at the SLC45A3 and ELK4 loci by 1M Affy 6.0 SNP Chips
Chimera classification system

**Class I: Inter-Chromosomal Translocation**

**Class II: Inter-Chromosomal Complex Rearrangements**

**Class III: Intra-Chromosomal Deletion**

**Class IV: Intra-Chromosomal Complex Rearrangements**

**Class V: Read-throughs**
Paired-end transcriptome strategy for chimera discovery
Comprehensiveness of paired-end sequencing
Broadly Expressed Vs. Restricted Chimeras

Broadly expressed chimeras
0% Inter-/Intra-chromosomal chimeras
100% Adjacent genes

Top ranking restricted chimeras
92.3% Inter-/Intra-chromosomal chimeras
7.7% Adjacent genes
Additional Classes of Chimeras Easily Observed Using Paired-End Sequencing
Conclusions-2

• “Re-discovered” recurrent gene fusions using both a hematological (BCR-ABL) and solid tumor (TMPRSS2-ERG) model

• Nominated, and experimentally validated, > 50 novel chimeras in cancer cell lines and tumors

• Demonstrated cell line can harbor multiple gene fusions many of which are likely to be “private” non-specific fusions

• Identified SLC45A3-ELK4 as a recurrent prostate cancer mRNA chimera not attributable to a DNA aberration

• Established chimera classification system for future categorization of this important class of cancer-related mutations

• Paired-end sequencing offers greater dynamic range and comprehensive assessment of chimeric mRNAs in a sample

• Developed a robust pipeline for chimeras discovery using high throughput sequencing that could reveal ideal therapeutic targets in common epithelial tumors
Metabolomic Profiling of Cancer Progression

Gleevec (A drug that can block the BCR-ABL gene fusion)

- Gleevec is a drug that targets the BCR-ABL gene fusion (>10 yrs to develop drug)
- Gleevec induced dramatic responses in patients with CML, including complete remission
- Gleevec is now a first line treatment of CML
- Prototype of rationally designed cancer therapeutics
Next-gen Transcriptome Sequencing

1. mRNA Purification

2. mRNA Fragmentation

3. cDNA Sequencing Library

4. Transcriptome Sequence

<table>
<thead>
<tr>
<th>Sample</th>
<th>K562</th>
<th>VCaP</th>
<th>LNCaP</th>
<th>RWPE</th>
<th>VCaP-Met</th>
<th>Met 3</th>
<th>Met 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total reads (millions)</td>
<td>66.9</td>
<td>76.4</td>
<td>57.3</td>
<td>71.9</td>
<td>14</td>
<td>35</td>
<td>9.24</td>
</tr>
<tr>
<td>Pass filter (millions)*</td>
<td>38.3</td>
<td>57.25%</td>
<td>40.3</td>
<td>52.75%</td>
<td>35.3</td>
<td>61.61%</td>
<td>44.8</td>
</tr>
<tr>
<td>Non-mapping reads (millions)**</td>
<td>2.08</td>
<td>5.43%</td>
<td>12.69</td>
<td>31.49%</td>
<td>1.59</td>
<td>4.50%</td>
<td>1.77</td>
</tr>
<tr>
<td>Redundantly mapping reads (millions)**</td>
<td>1.42</td>
<td>3.71%</td>
<td>1.08</td>
<td>2.68%</td>
<td>1.23</td>
<td>3.48%</td>
<td>1.74</td>
</tr>
<tr>
<td>Best hit uniquely maps (millions)**</td>
<td>19.86</td>
<td>51.85%</td>
<td>15.48</td>
<td>38.41%</td>
<td>19.34</td>
<td>54.79%</td>
<td>26.13</td>
</tr>
<tr>
<td>Mitochondrial reads (millions)**</td>
<td>1.89</td>
<td>4.93%</td>
<td>1.72</td>
<td>4.27%</td>
<td>3.19</td>
<td>9.04%</td>
<td>2.8</td>
</tr>
<tr>
<td>Ribosomal reads (millions)**</td>
<td>13.09</td>
<td>34.18%</td>
<td>9.35</td>
<td>23.20%</td>
<td>10</td>
<td>28.33%</td>
<td>12.34</td>
</tr>
</tbody>
</table>

* Percentages relative to total reads
** Percentages relative to reads passing filter
Breakdown of Paired End Mappings

- ~20 & 17 million K562 and VCaP PE reads, respectively
- Majority map to same transcript
- Higher percentage of non-mapping in VCaP likely due to viral sequences
- ~1% of each cell line are categorized as chimera candidates