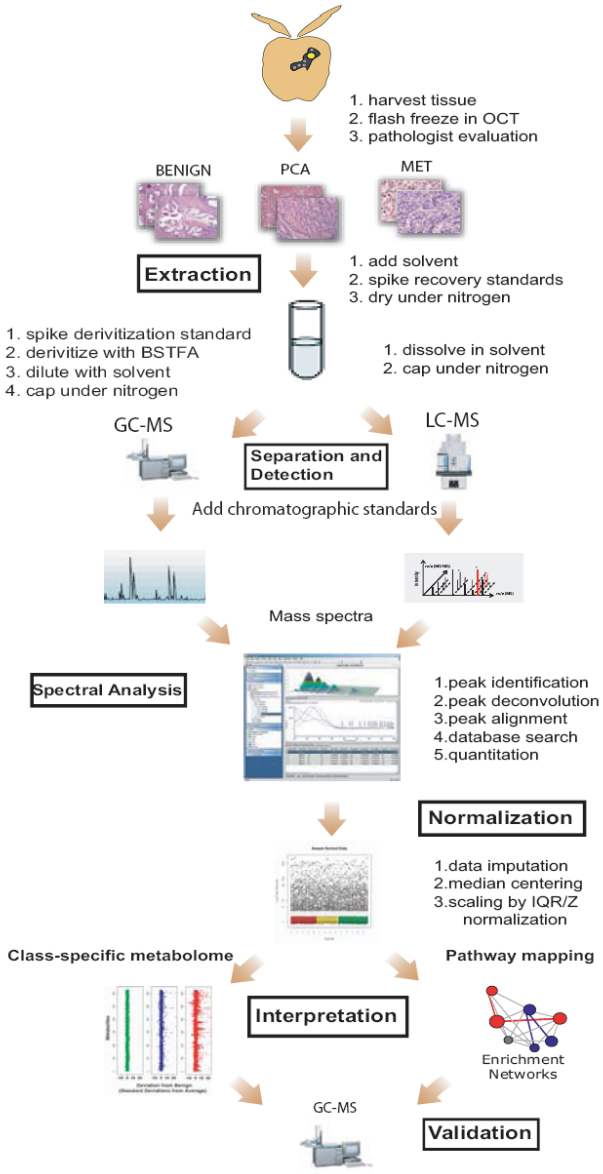
# The Next Generation of Gene Fusion Discovery in Cancer

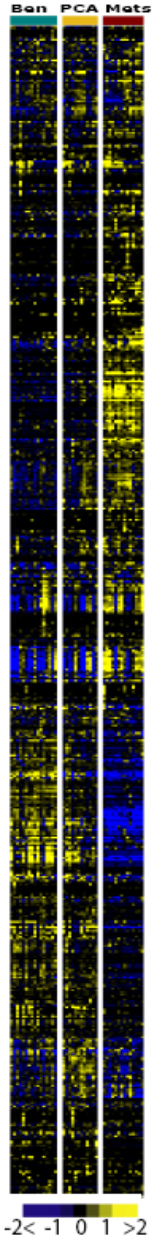
**Arul M. Chinnaiyan, M.D., Ph.D.**
**Departments of Pathology and Urology**
**Howard Hughes Medical Institute**

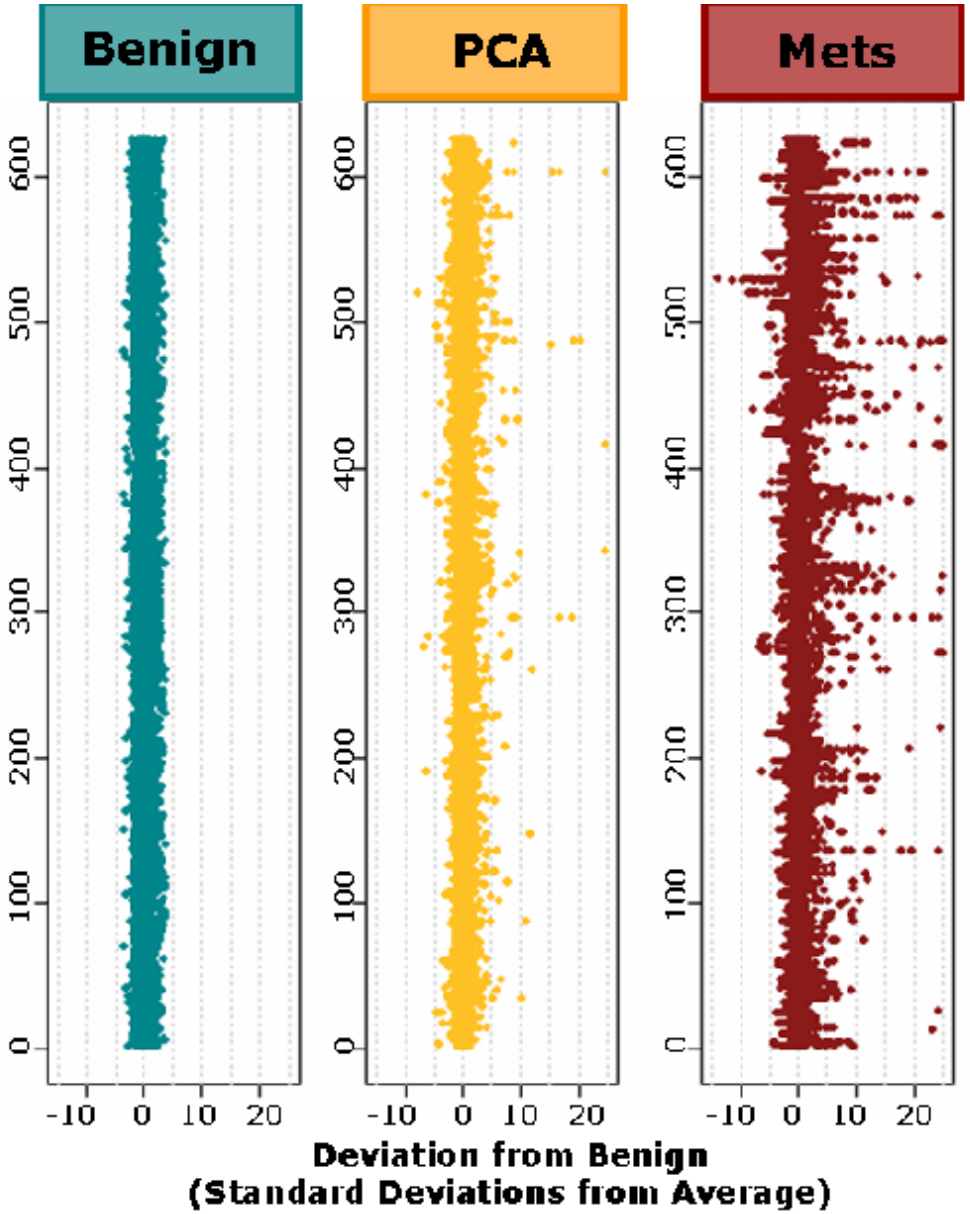

# Areas where the Cancer DBP will interact with the NCIBI

- Bioinformatics related to metabolomics
- Integrative analysis across molecular alterations
- Bioinformatics related to Next Generation sequencing

# Metabolomic Profiling of Cancer Progression

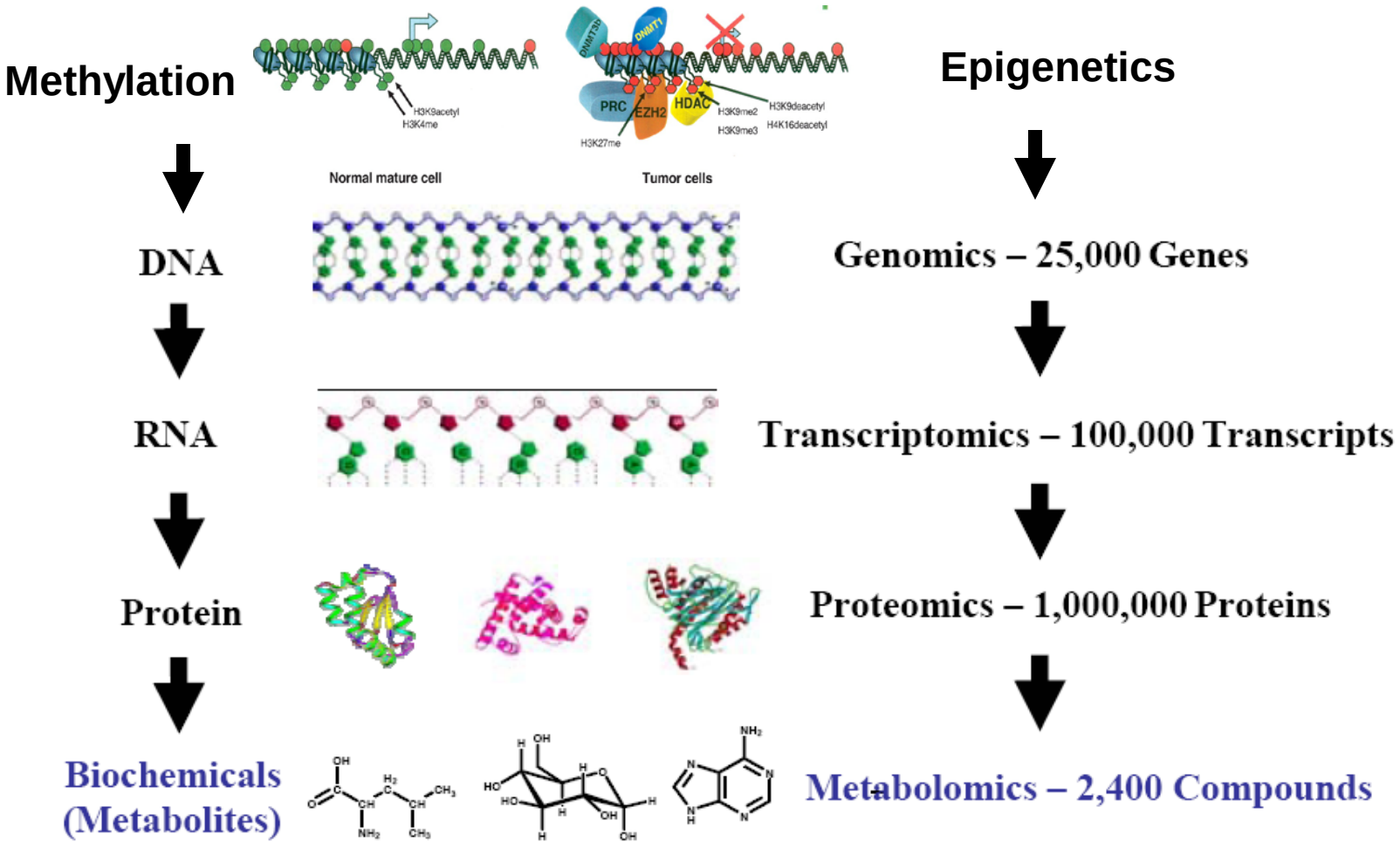



Sreekumar et al, *Nature* 2009

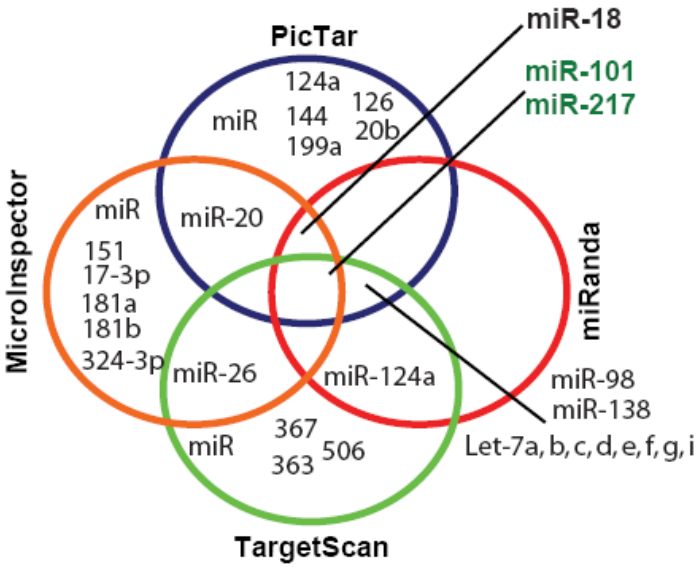
# Complexity of the -omics

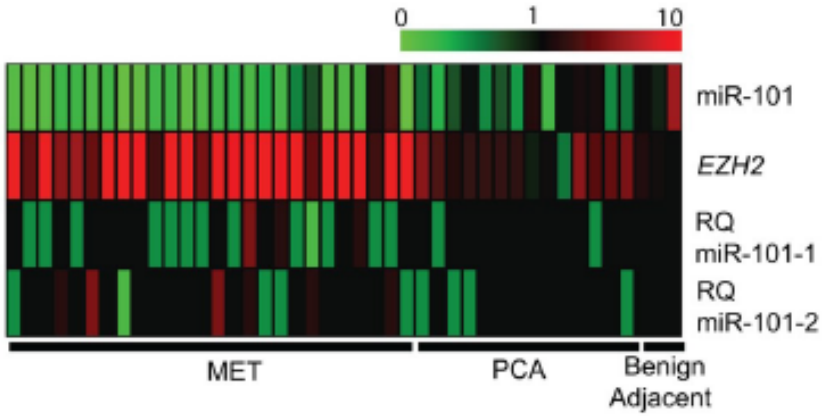


Sciencexpress Report

# Genomic Loss of microRNA-101 Leads to Overexpression of Histone Methyltransferase EZH2 in Cancer

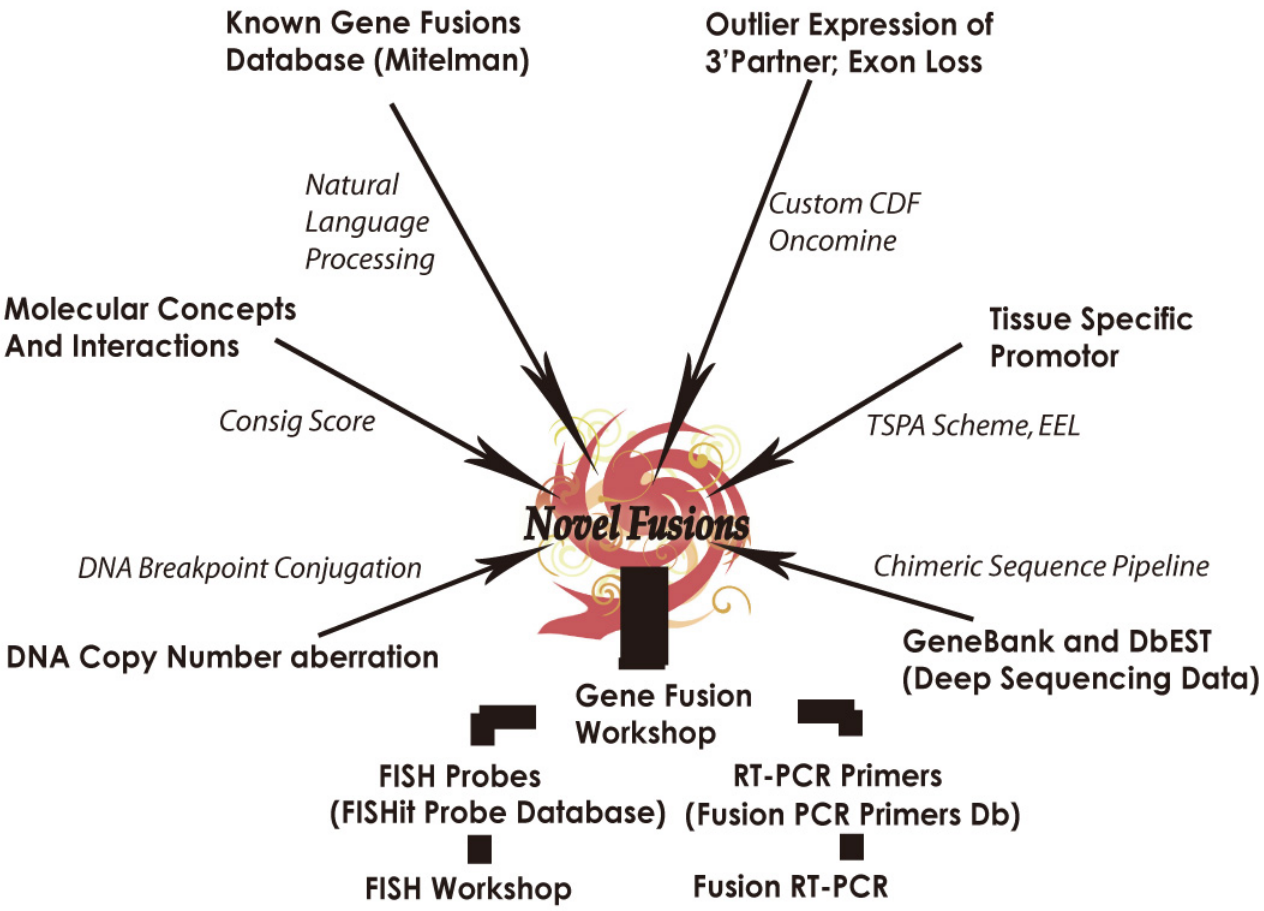
Sooryanarayana Varambally,[1,3,6]* Qi Cao,[1,3]* Ram-Shankar Mani,[1,3] Sunita Shankar,[1,3] Xiaosong Wang,[1,3] Bushra Ateeq,[1,3] Bharathi Laxman,[1,3] Xuhong Cao,[1,2] Xiaojun Jing,[1,3] Kalpana Ramnarayanan,[7] J. Chad Brenner,[1,3,5] Jindan Yu,[1,3] Jung H. Kim,[1,3] Bo Han,[1,3] Patrick Tan,[7,8] Chandan Kumar-Sinha,[1,3] Robert J. Lonigro,[1,6] Nallasivam Palanisamy,[1,3,7] Christopher A. Maher,[1,3] Arul M. Chinnaiyan[1,2,3,4,5,6]†




Nov 2008

# *The integrative model for translation of bio-data into novel gene fusions*





***Xiaosong Wang***

*Arul M. Chinnaiyan*

*Saravana M Dhanasekaran*

*John R. Prensner*

*Bo Han*

*Nallasivam Palanisamy*

*Sunita Shankar*

***Xiaosong Wang***

*Gilbert S. Omenn*

*Maureen A. Sartor*

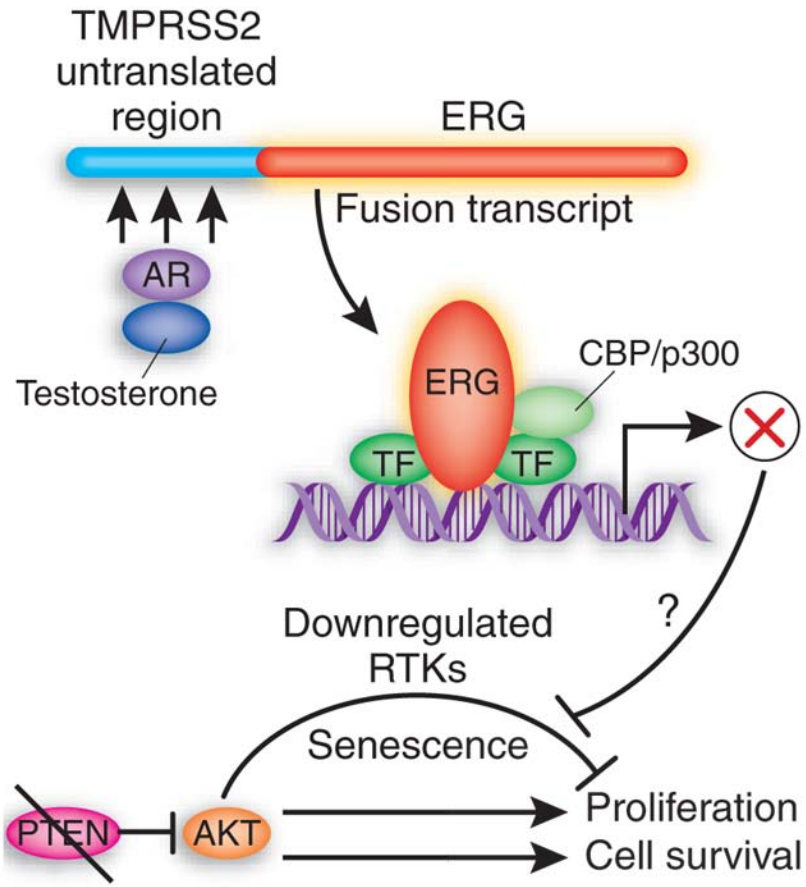
*James Cavocoli*

*Fan Meng*

*Alex Ade*

*Lihshwu Ke*

- Transcriptome
  - Gene expression
  - Alternative Splice (AS) Variants
  - Mutation discovery
  - Gene Fusion Discovery
- Digital Gene Expression
  - Enumerate gene expression
- Methylation
- Small RNA
  - miRNA profiling
  - Discovery
- ChIP-Seq
  - Binding site detection



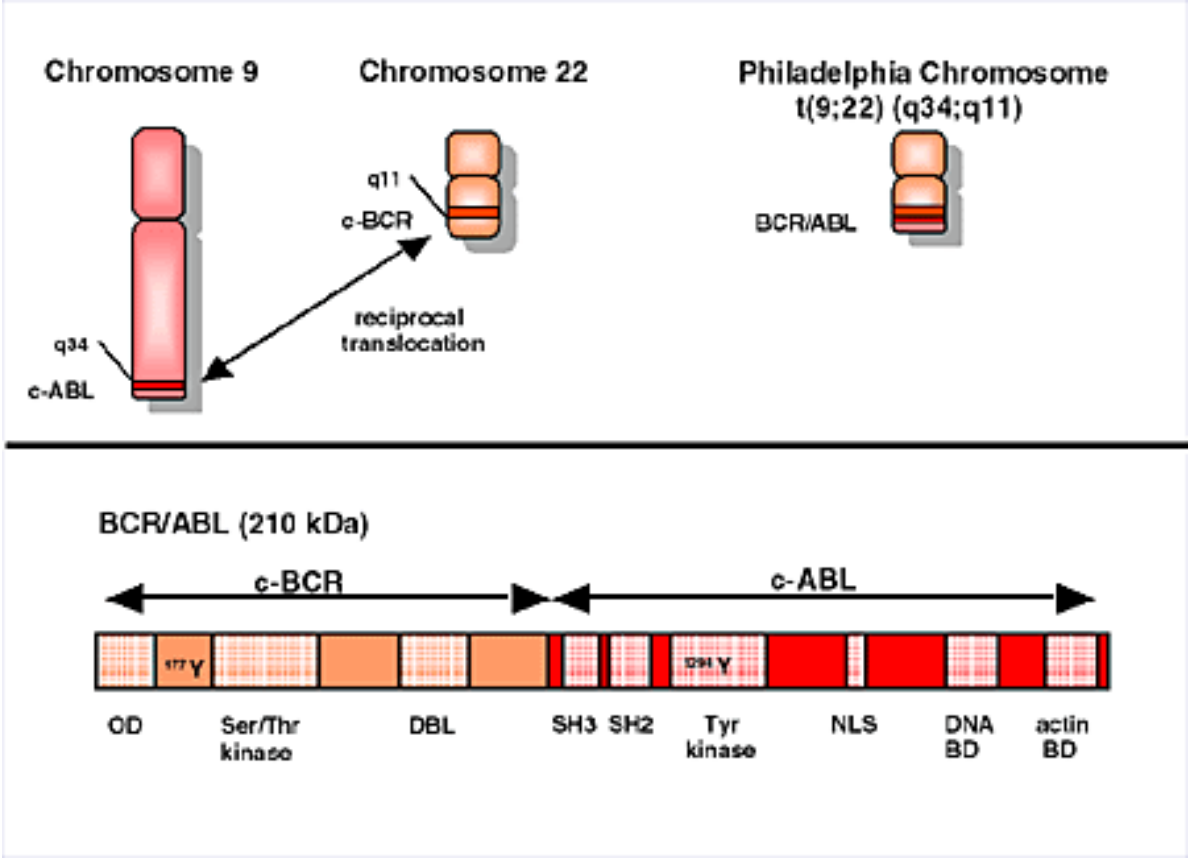

# Breaking the rules of cancer

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.**

University of Michigan
Medical School

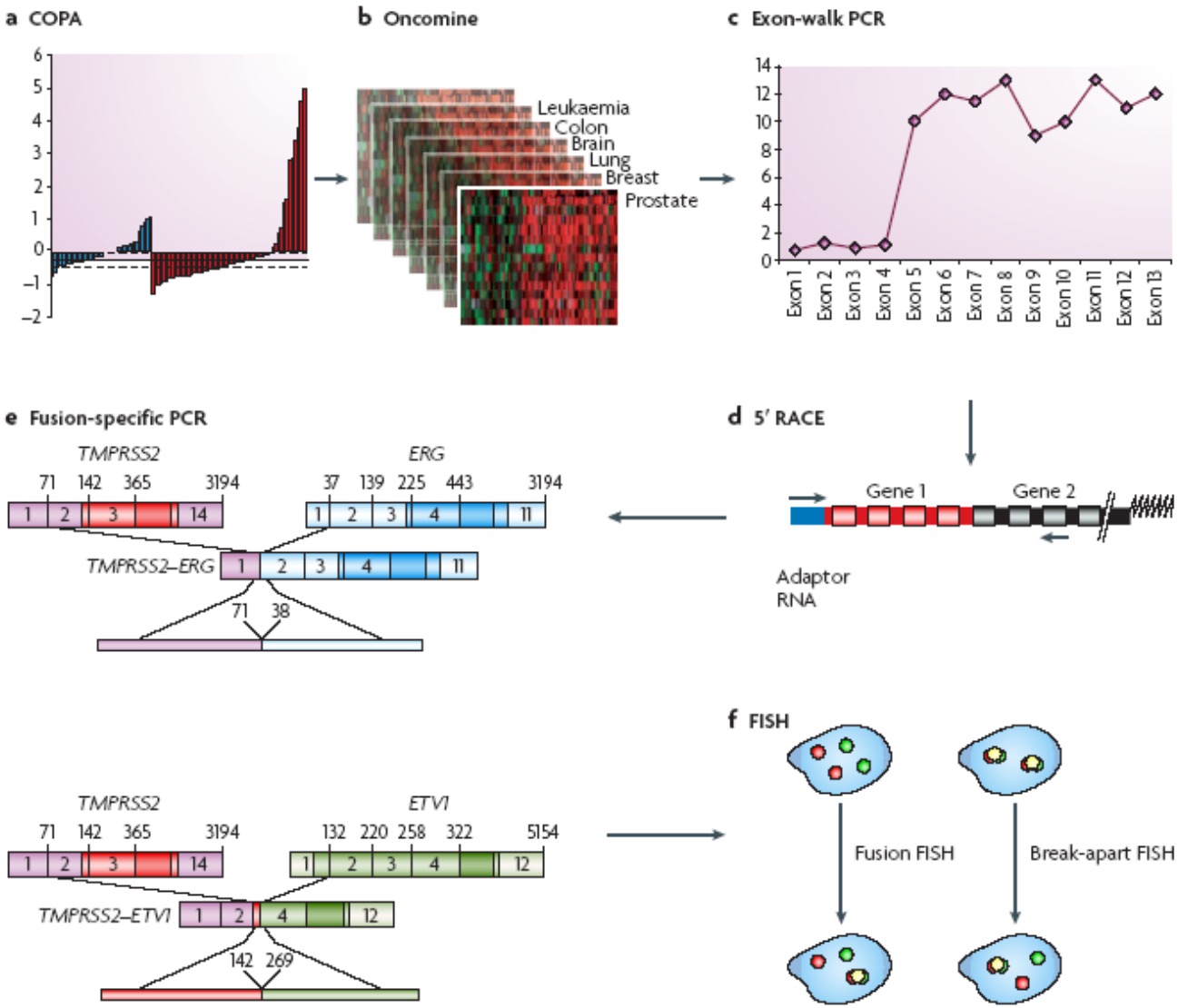
# BCR-ABL Gene Fusion in CML (a type of leukemia)




http://www.cmlsupport.com/bcrabl.gif

University of Michigan
Medical School

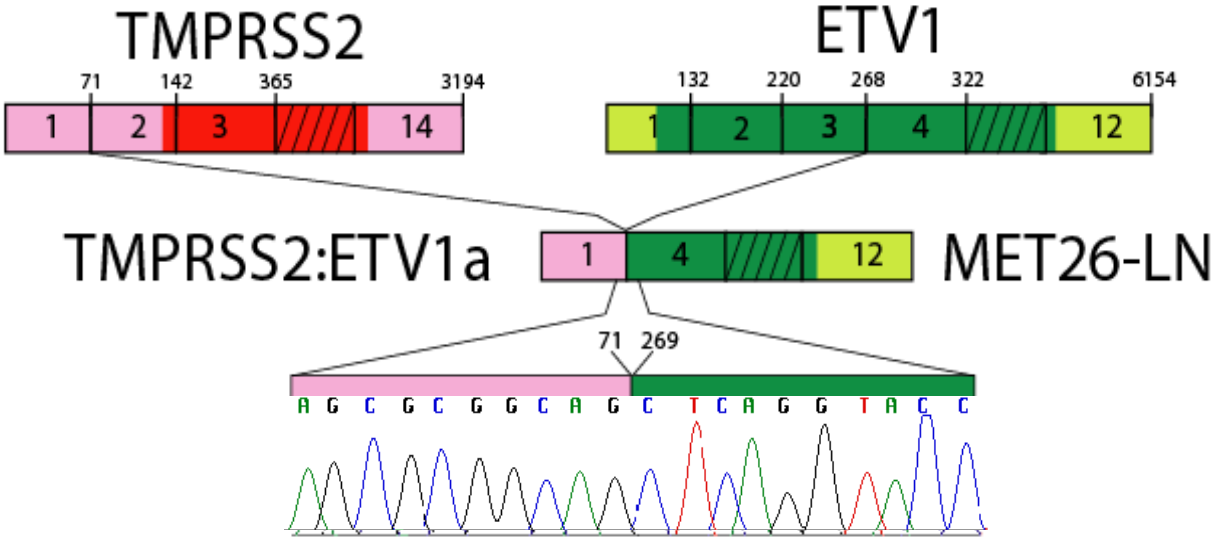
# A Computational Approach Leads to the Discovery of Gene Fusions in Prostate Cancer

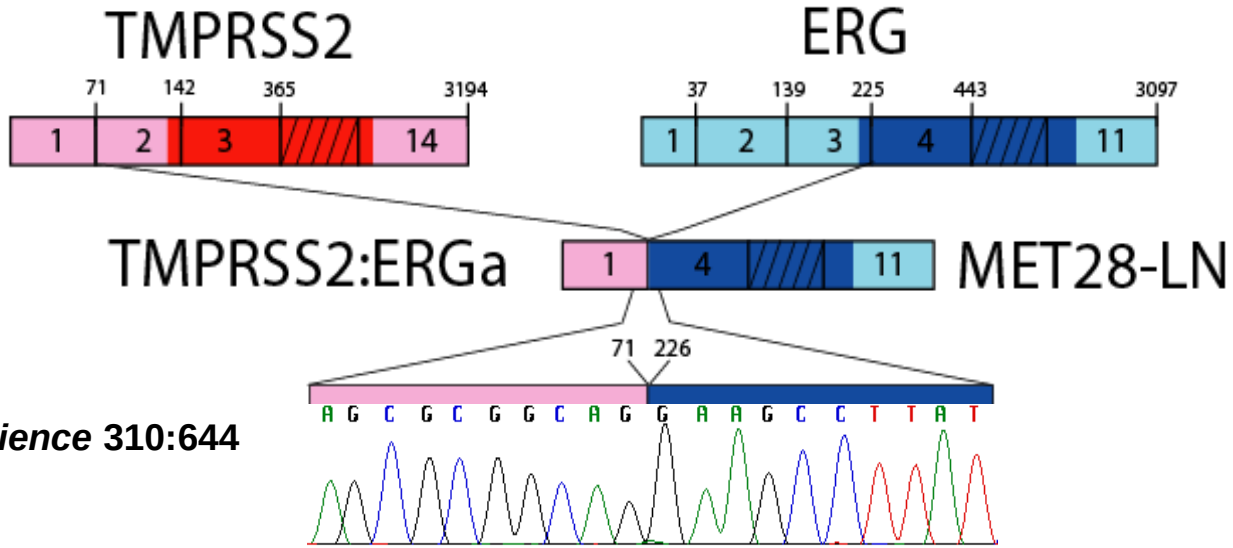



Kumar et al *Nature Cancer Reviews*, 2008

University of Michigan Medical School

# TMPRSS2-ETS Gene Fusions in Prostate Cancer




Tomlins et al *Science* 310:644

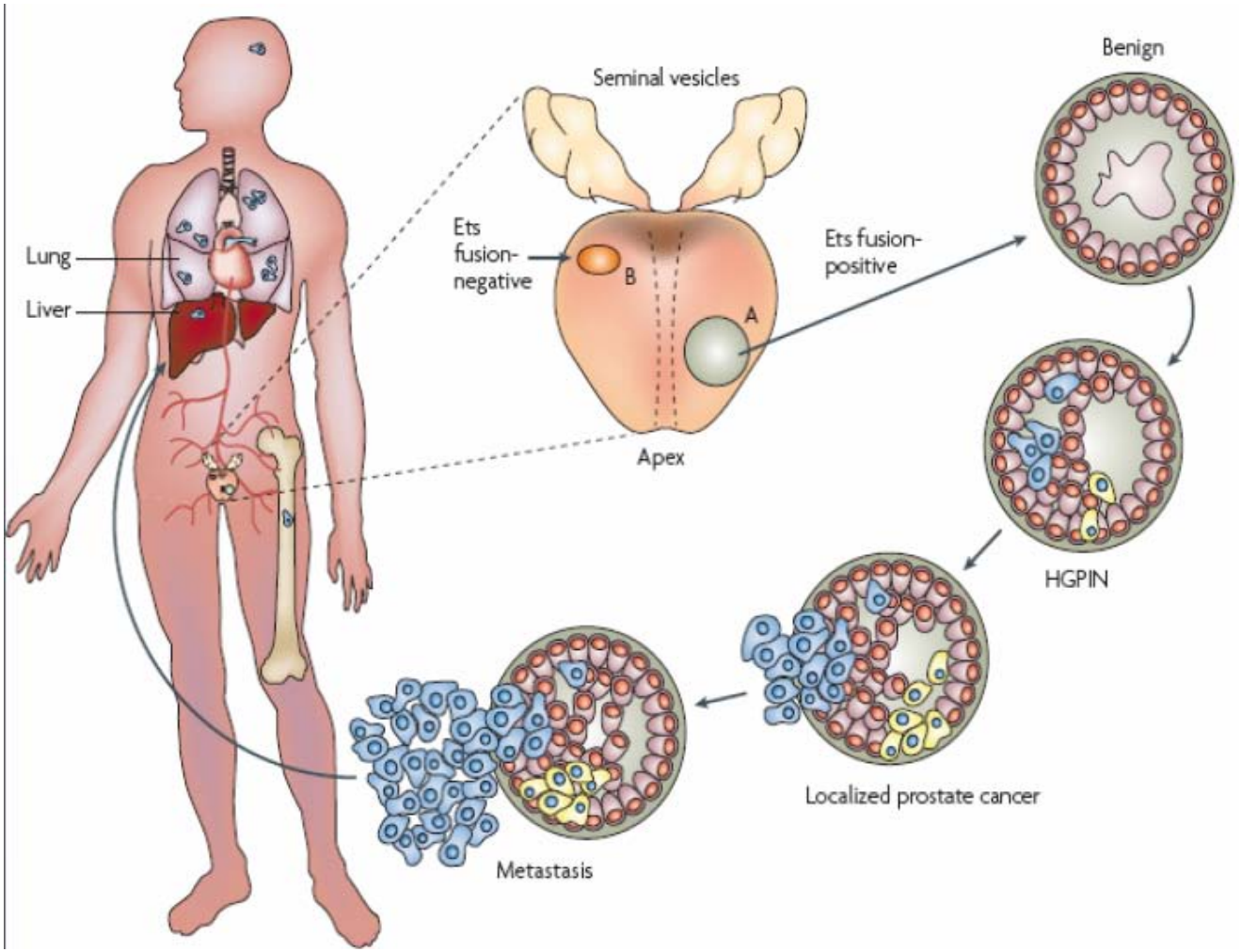
University of Michigan Medical School

# Confirmation of High Prevalence of Gene Fusions (40-80%)


- University of Michigan
- Harvard
- U of Toronto
- Baylor College of Medicine
- Erasmus, The Netherlands
- Finland
- Sweden
- UK
- Portugal
- Gen-Probe, Inc.
- Others

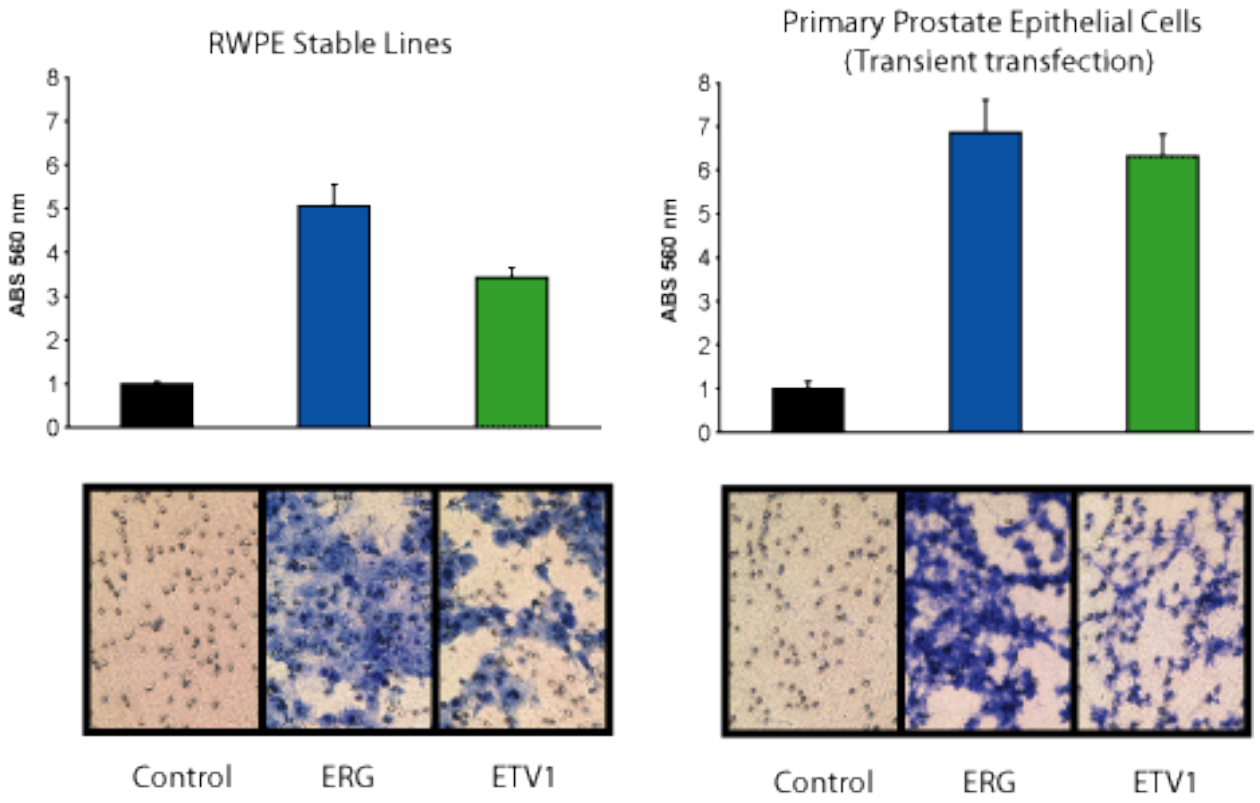
University of Michigan
Medical School

# 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**

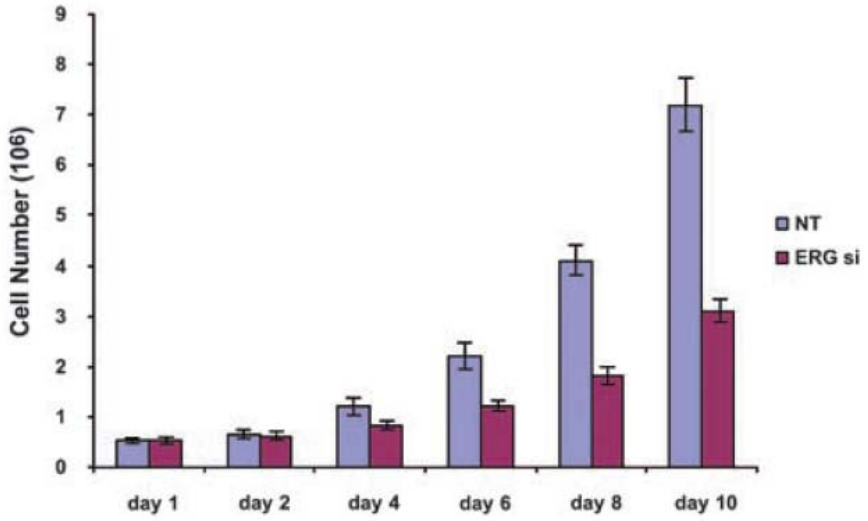
# ETS Gene Fusion Products Induce Cell Invasion



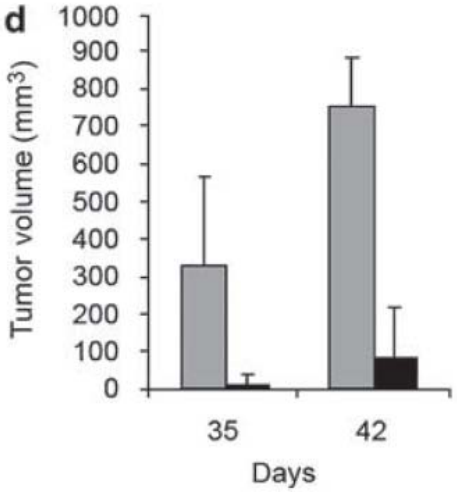

University of Michigan
Medical School

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

**VCaP cells (TMPRSS2-ERG +)**

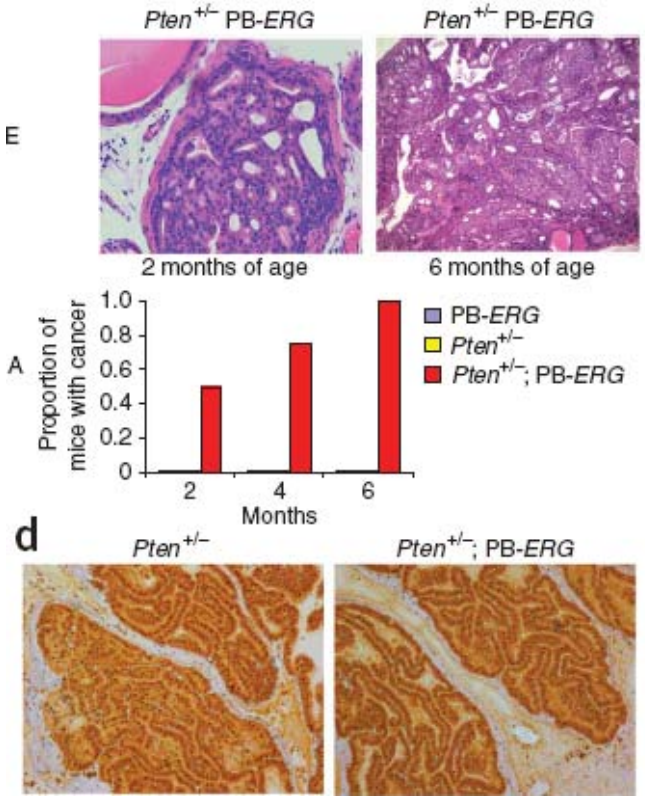



**VCaP Xenografts**




**Shiv Srivastava et al**
***Oncogene* (2008) 27:5348**

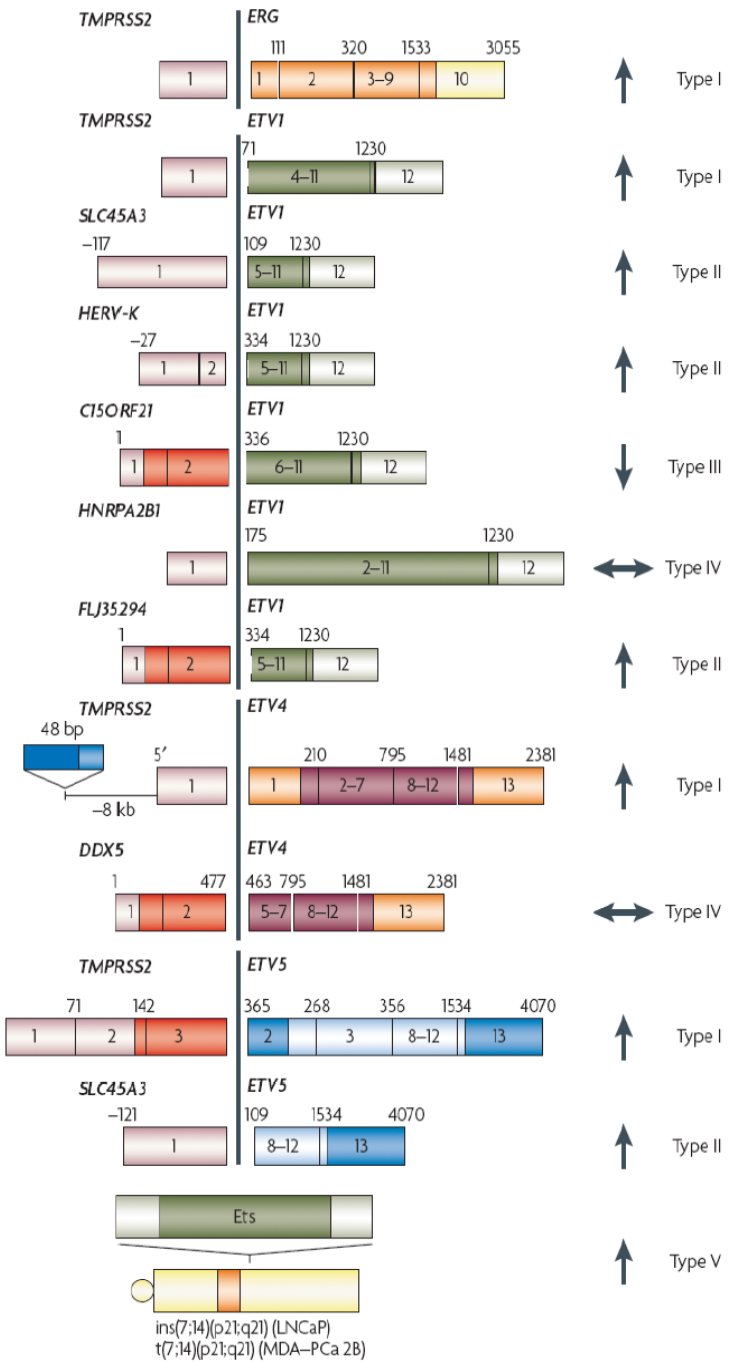
nature genetics




# Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate

Brett S Carver[1,2], Jennifer Tran[1], Anuradha Gopalan[3], Zhenbang Chen[1,4], Safa Shaikh[2], Arkaitz Carracedo[1,4], Andrea Alimonti[1,4], Caterina Nardella[1,4], Shohreh Varmeh[1,4], Peter T Scardino[2], Carlos Cordon-Cardo[5], William Gerald[3] & Pier Paolo Pandolfi[1,3,4]

# 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

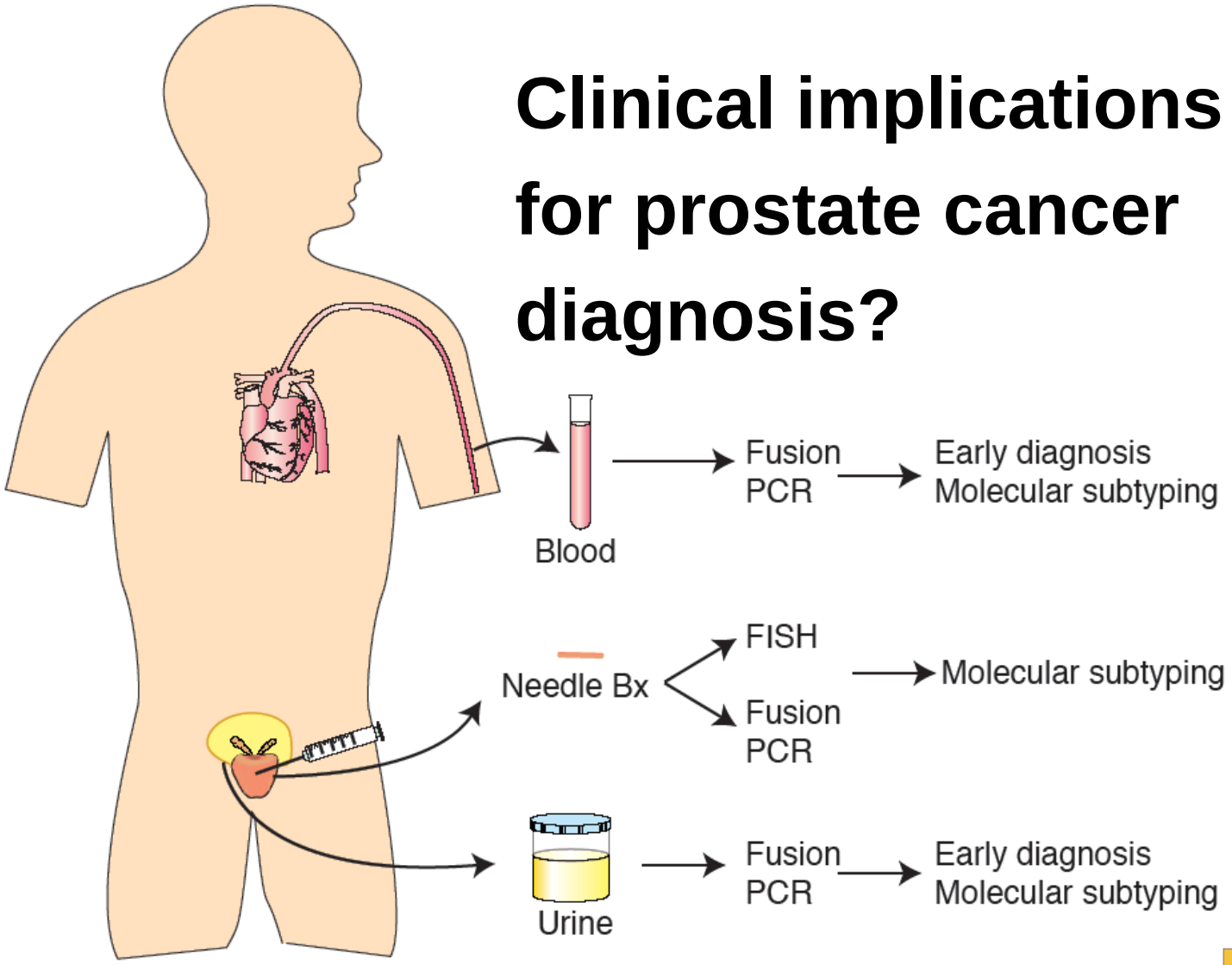
TMPRSS2-ERG with deletion (50-60%)
TMPRSS2-ERG without deletion (40-50%)
About 15 variant fusion transcripts

~5-10% ETV1Fusions
~1% ETV4 Fusions
~1% ETV5 Fusions

- **~30% Negative for ETS gene fusions**

University of Michigan Medical School

# Clinical implications for prostate cancer diagnosis?

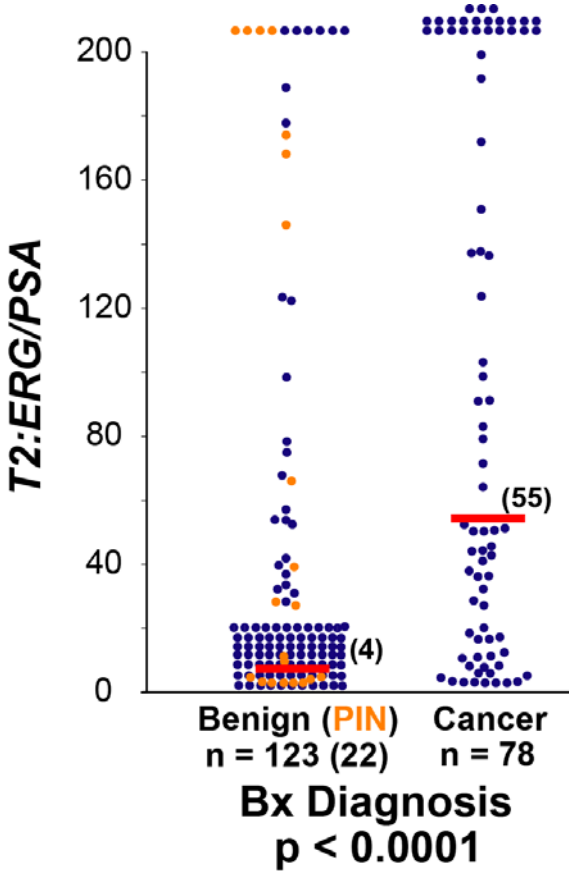


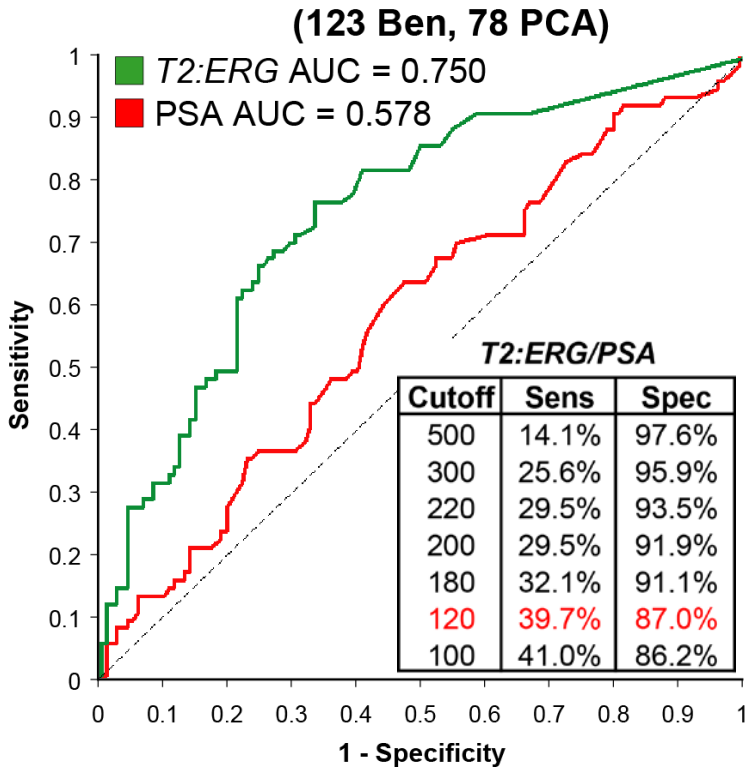

University of Michigan
Medical School

# 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



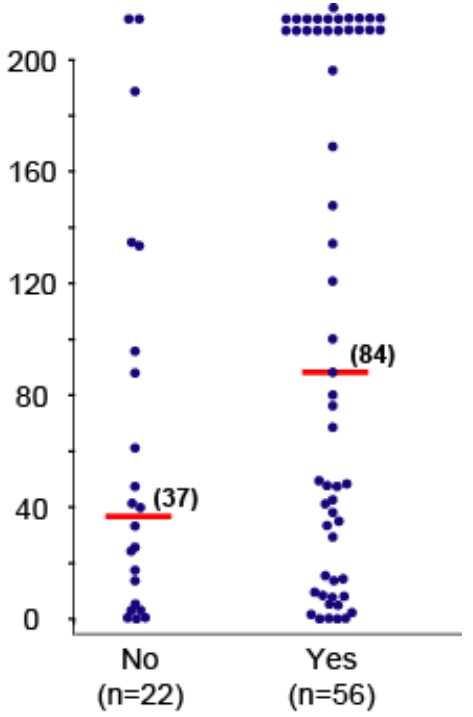




T2:ERG/PSA

| Cutoff | Sens | Spec |
|---|---|---|
| 500 | 14.1% | 97.6% |
| 300 | 25.6% | 95.9% |
| 220 | 29.5% | 93.5% |
| 200 | 29.5% | 91.9% |
| 180 | 32.1% | 91.1% |
| 120 | 39.7% | 87.0% |
| 100 | 41.0% | 86.2% |

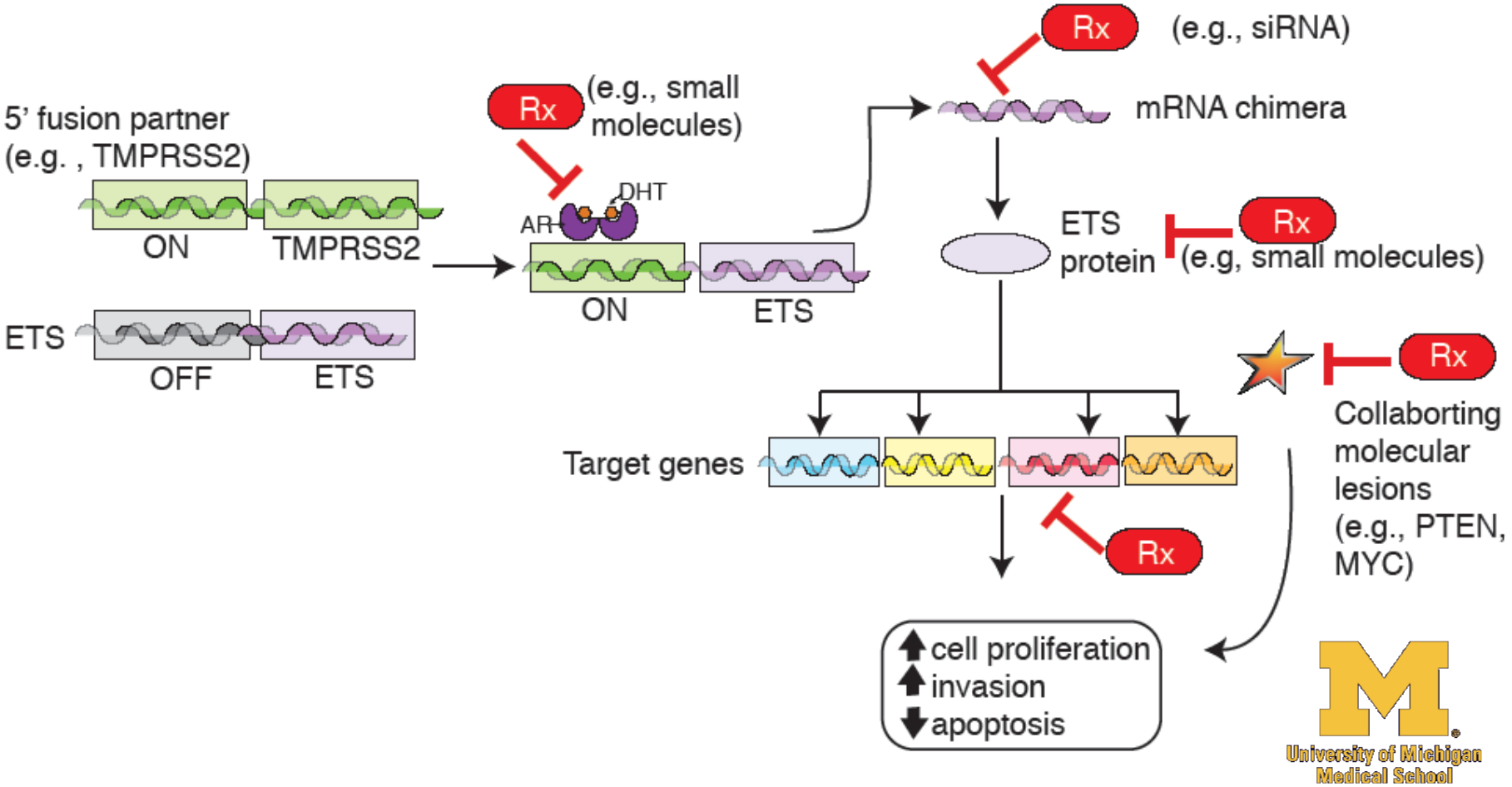
Serum PSA

| Cutoff | Sens | Spec |
|---|---|---|
| 2.5 | 92.3% | 16.3% |
| 4 | 76.9% | 33.3% |
| 10 | 15.4% | 87.0% |

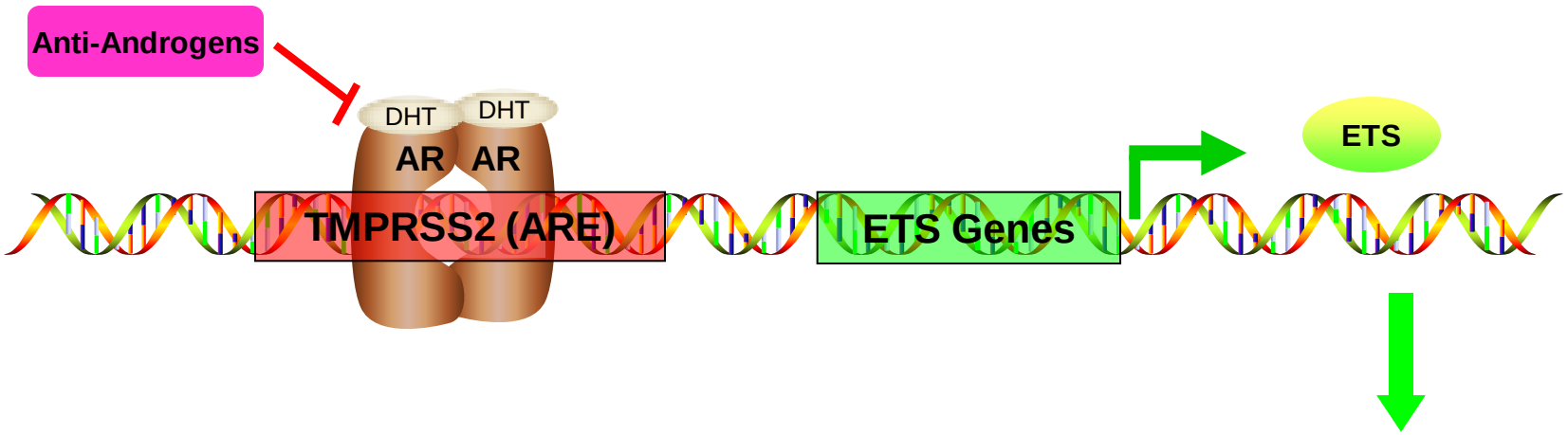



**Similar results (cancer vs. benign, significant vs. insignificant) in SDVA/UL cohort**

# Prostate cancer therapy: How do we target the prostate cancer gene fusions?



# A Molecular Basis for Prostate Cancer



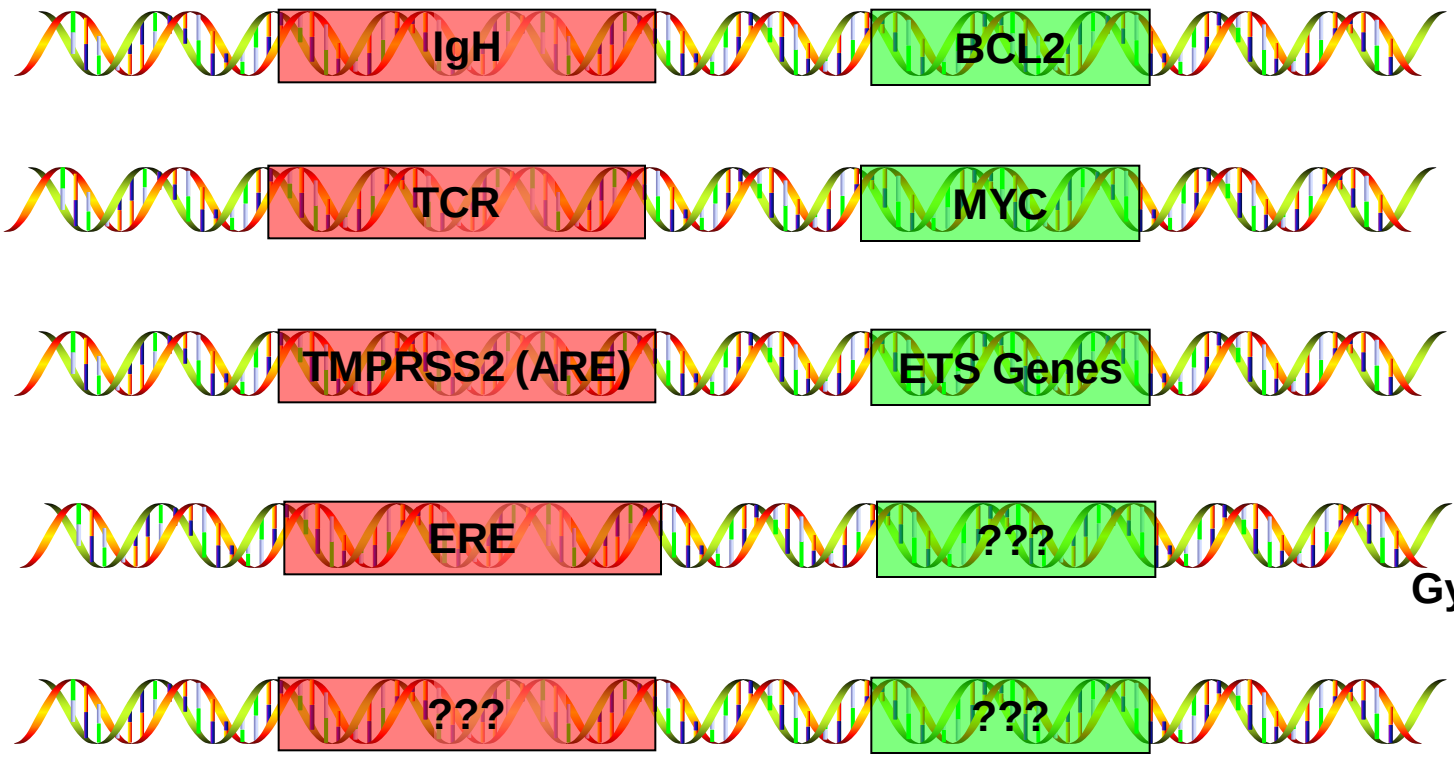

**ETS Target Genes**

- Cell growth
- Invasion/Metastasis
- Cell Survival

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

University of Michigan Medical School

# Subversion of Tissue-Specific Promoters/Enhancers to Cause Cancer




B cell lymphomas

T cell lymphomas

Prostate Cancer

Breast Cancer?
Gynecologic Cancer?

Others?
(e.g., Lung, Colon, Ovarian, Liver)

University of Michigan Medical School

doi:10.1038/nature05945

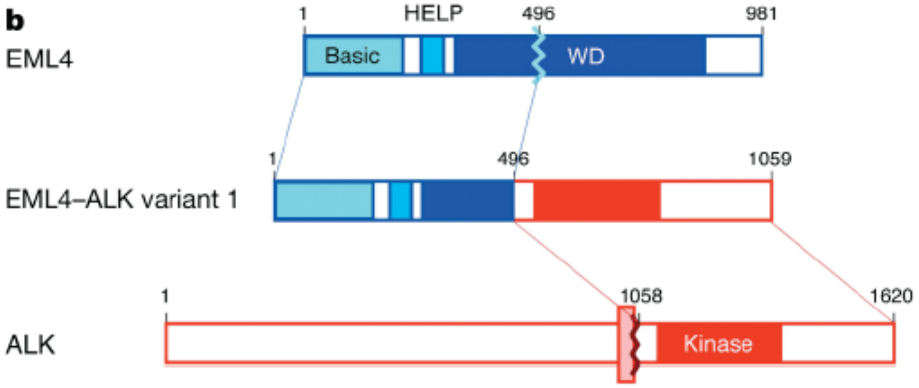
nature

ARTICLES

# Identification of the transforming *EML4–ALK* fusion gene in non-small-cell lung cancer

Manabu Soda[1,2], Young Lim Choi[1], Munehiro Enomoto[1,2], Shuji Takada[1], Yoshihiro Yamashita[1], Shunpei Ishikawa[5], Shin-ichiro Fujiwara[1], Hideki Watanabe[1], Kentaro Kurashina[1], Hisashi Hatanaka[1], Masashi Bando[2], Shoji Ohno[2], Yuichi Ishikawa[6], Hiroyuki Aburatani[5,7], Toshiro Niki[3], Yasunori Sohara[4], Yukihiko Sugiyama[2] & Hiroyuki Mano[1,7]

[1]Division of Functional Genomics, [2]Division of Pulmonary Medicine, [3]Department of Pathology, and [4]Division of General Thoracic Surgery, Jichi Medical University, Tochigi 329-0498, Japan. [5]Research Center for Advanced Science and Technology, University of Tokyo, Tokyo 153-8904, Japan. [6]Department of Pathology, The Cancer Institute, Japanese Foundation for Cancer Research, Tokyo 135-8550, Japan. [7]Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency, Saitama 332-0012, Japan.



# Conclusions-1

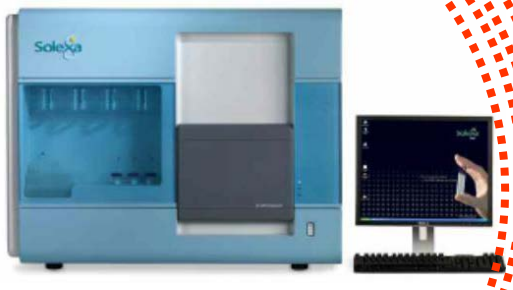
- **Majority of prostate cancers have a gene fusion/translocation**
- **ETS rearrangements in prostate cancer are the most common gene fusion in cancer**
- **We propose that gene fusions of *prostate-active genomic regulatory regions* to *oncogenic factors* is the basis for the development of prostate 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

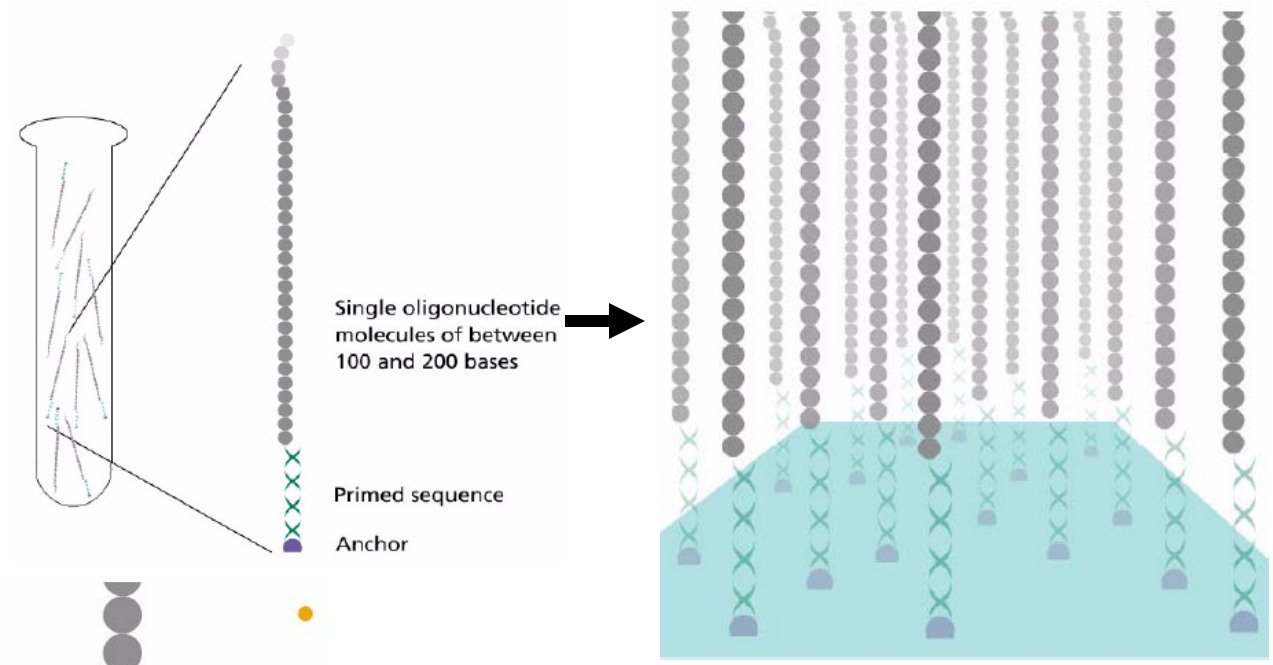
Roche / 454
Genome Sequencer FLX
100 Mb /run



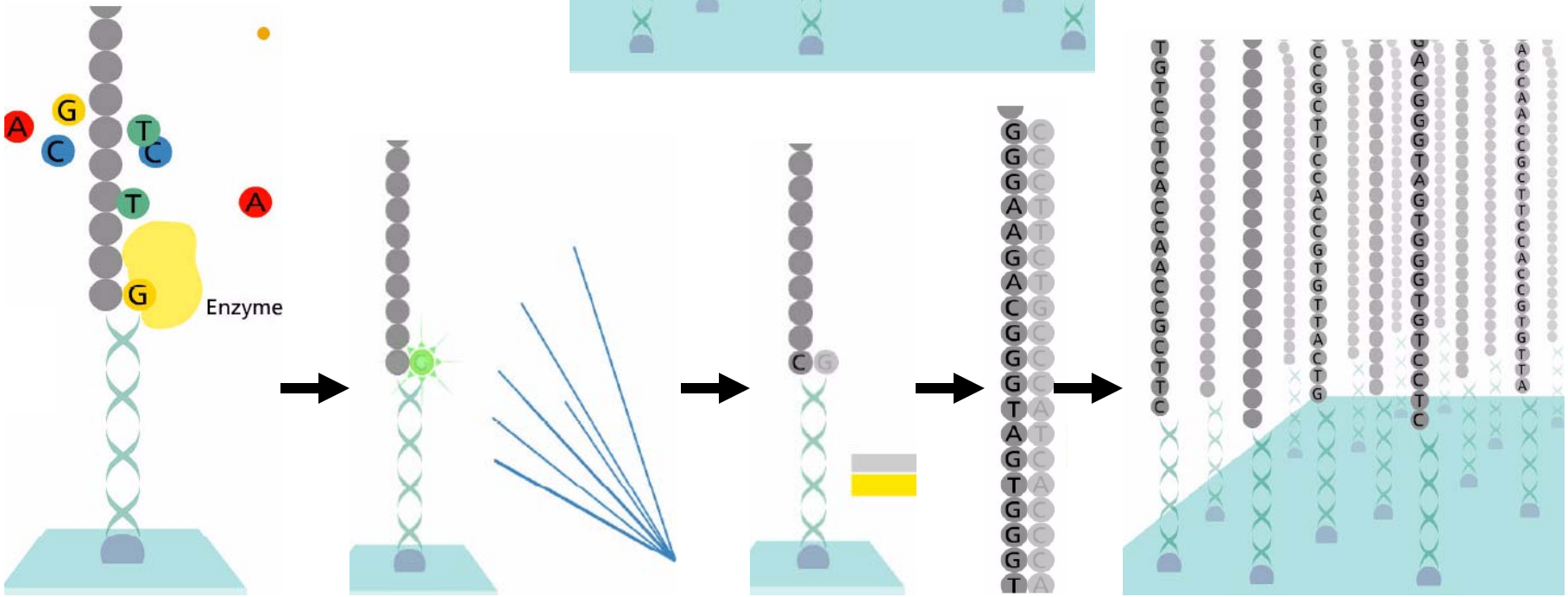
Illumina / Solexa
Genetic Analyzer
2000 Mb /run

Reference Genome
Reference Genome

# Massively Parallel, High throughput, Next-Gen sequencing

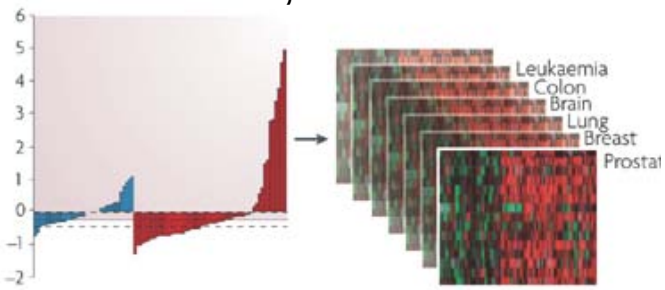



**Adapted from:**
**Dr Simon Bennett**
**www.solexa.com**

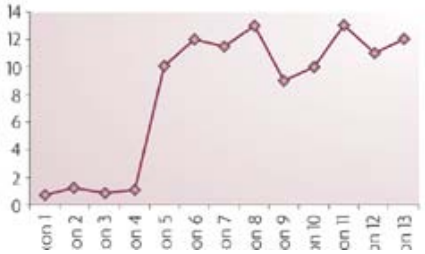


# Next Generation Sequencing Fusion Pipeline

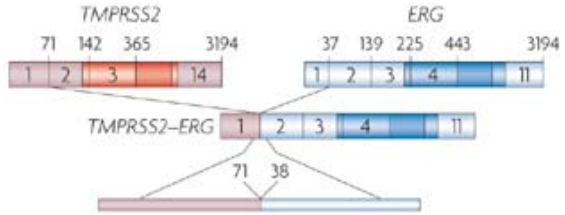
COPA ; Oncomine



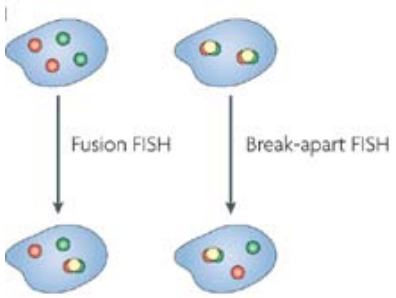

Exon-walking (qPCR)



Fusion-specific PCR




FISH

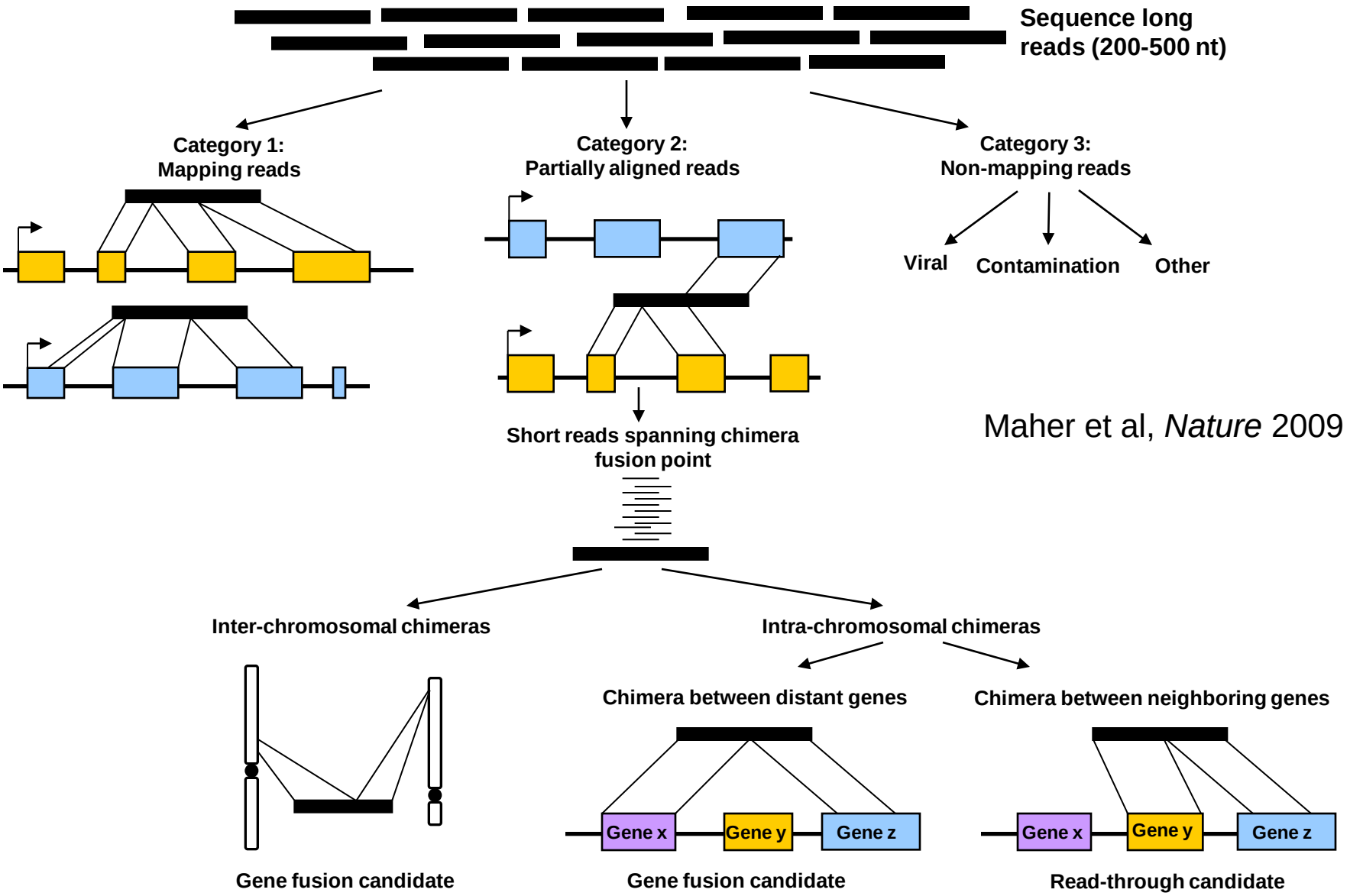



**(Kumar-Sinha, et al., Nature Cancer Reviews 2008)**

# Next Generation Transcriptome Sequencing for gene fusion discovery

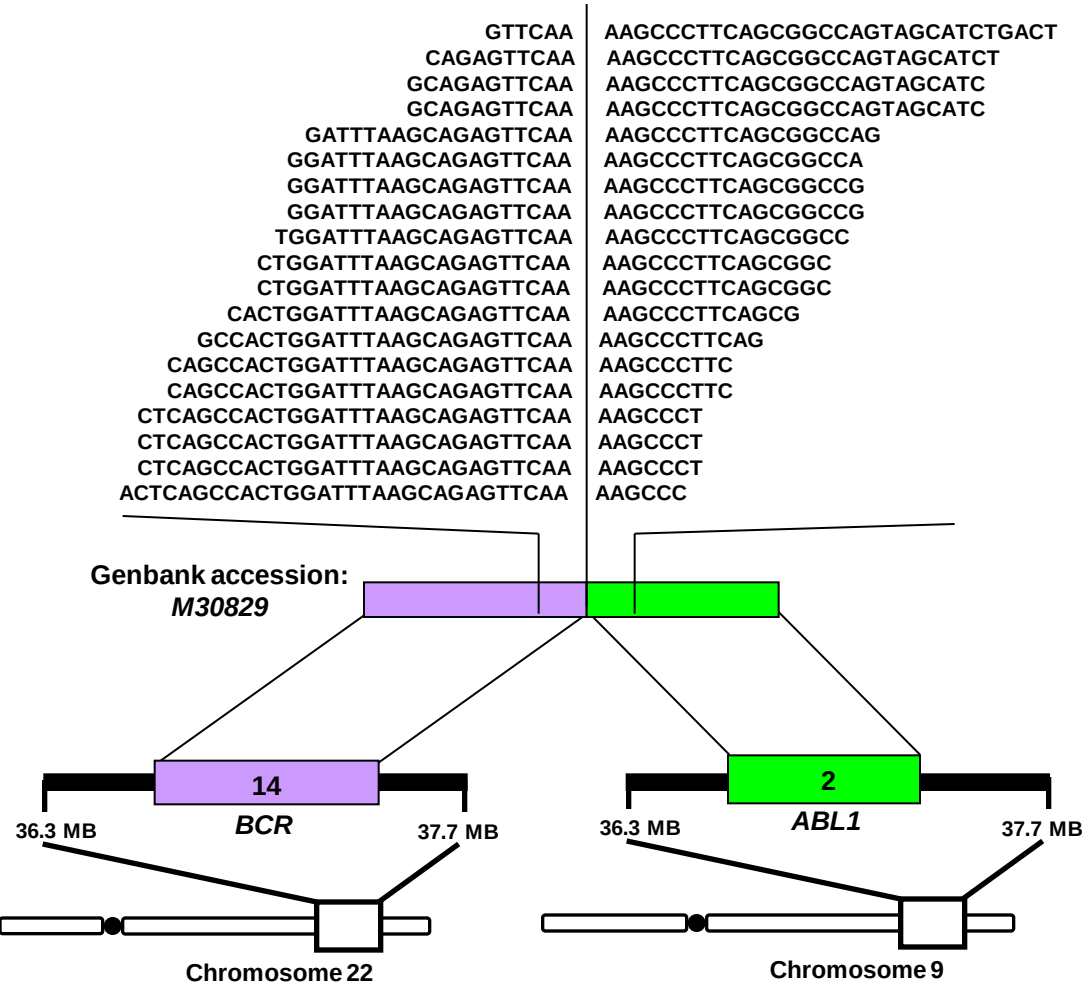
- Circumvents limitations introduced by cytogenetic techniques
- Increased sensitivity over microarrays
- Unbiased view of expressed portion of the genome
- Requires less resources than genomic sequencing, while focusing on aberrations are primarily expected to have functional consequences in carcinogenesis
- 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



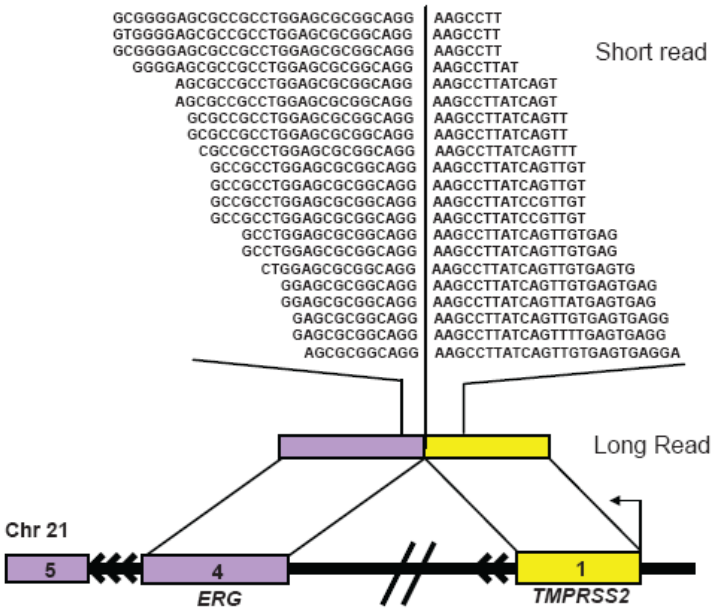

Maher et al, *Nature* 2009

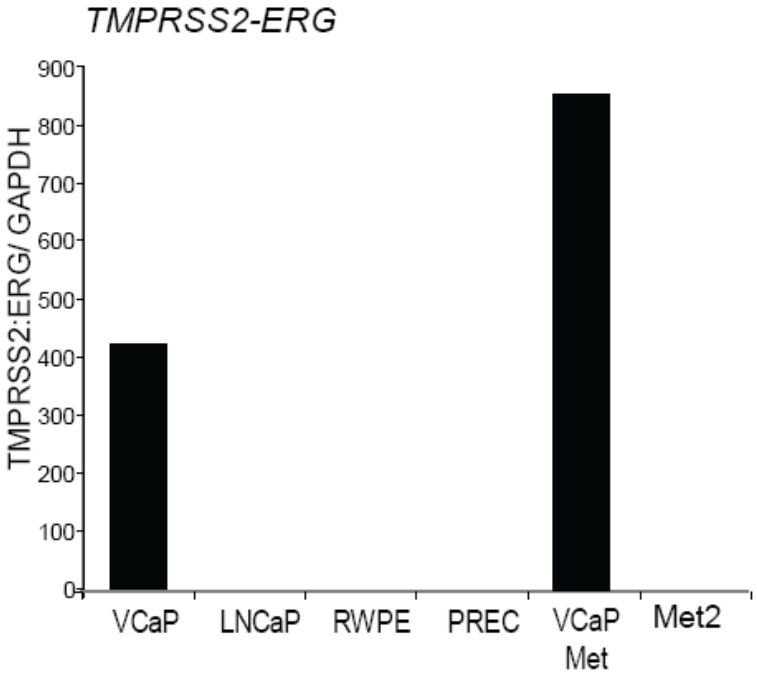
# “Re-discovery” of *BCR-ABL1*




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

# “Re-discovery” of *TMPRSS2-ERG*



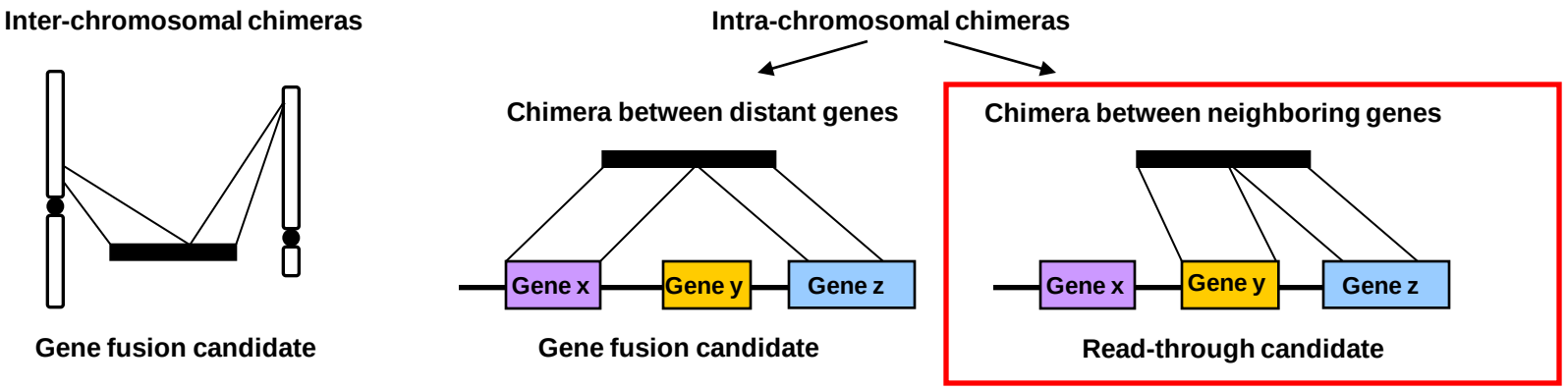

- ~76 million 36-mer sequence reads from 8 lanes
- 21 chimeras spanning *TMPRSS2-ERG* fusion boundary in VCaP cells

**Supplementary Table 7. Overall summary of validated chimeras.** In-frame chimeras are denoted with an asterik.

| Chimera | Chimera Class | Location | 5' Gene | Location | 3' Gene | Validated in | Validated by |
|---|---|---|---|---|---|---|---|
| *BCR-ABL1* | Class I: Translocation | 22q11.23 | BCR, breakpoint cluster region | 9q34.1 | ABL1, c-abl oncogene 1, receptor tyrosine kinase | K562 | Short read, qRT-PCR |
| *MRPS10-HPR* | Class I: Translocation | 6p21.1 | MRPS10, mitochondrial ribosomal protein S10 | 16q22.1 | HPR, haptoglobin-related protein | LNCaP | Long read, Short read, qRT-PCR, FISH |
| *MIPOL1-DGKB* | Class II: Inter-chromosomal complex | 14q13.3-q21.1 | MIPOL1, mirror-image polydactyly 1 | 7p21.2 | DGKB, diacylglycerol kinase, beta 90kDa | LNCaP | Long read, Short read, qRT-PCR. FISH |
| *TMPRSS2-ERG** | Class III: Interstitial Deletion | 21q22.3 | TMPRSS2, transmembrane protease, serine 2 | 21q22.3 | ERG, v-ets erythroblastosis virus E26 oncogene homolog (avian) | VCaP, VCaP-Met | Long read, Short read, qRT-PCR, FISH |
| *USP10-ZDHHC7** | Class III: Interstitial Deletion | 16q24.1 | USP10, ubiquitin specific peptidase 10 | 16q24.1 | ZDHHC7, zinc finger, DHHC-type containing 7 | VCaP, VCaP-Met | Long read, Short read, qRT-PCR, aCGH |
| *STRN4-GPSN2** | Class IV: Intra-chromosomal complex | 19q13.2 | STRN4, striatin, calmodulin binding protein 4 | 19p13.12 | GPSN2, glycoprotein, synaptic 2 | Met-3 | Short read, qRT-PCR |
| *LMAN2-AP3S1* | Class IV: Intra-chromosomal complex | 5q35.3 | LMAN2 lectin, mannose-binding 2 | 5q22 | AP3S1, adaptor-related protein complex 3, sigma 1 | VCaP, VCaP-Met | Short read, qRT-PCR |
| *HJURP-EIF4E2** | Class IV: Intra-chromosomal complex | 2q37.1 | HJURP, Holliday junction recognition protein | 2q37.1 | EIF4E2, eukaryotic translation initiation factor 4E family member 2 | VCaP, VCaP-Met | Long read, Short read, qRT-PCR, FISH |
| *INPP4A-HJURP** | Class II: Intra-chromosomal complex | 2q11.2 | INPP4A, inositol polyphosphate-4-phosphatase, type I | 2q37.1 | HJURP, Holliday junction recognition protein | VCaP, VCaP-Met | Long read, Short read, qRT-PCR, FISH |
| *RC3H2-RGS3* | Class IV: Intra-chromosomal complex | 9q34 | RC3H2, ring finger and CCCH-type zinc finger domains 2 | 9q32 | RGS3, regulator of G-protein signaling 3 | VCaP, VCaP-Met | Short read, qRT-PCR |
| *ZNF649-ZNF577* | Class V: Read-through | 19q13.33 | ZNF649, zinc finger protein 649 | 19q13.33 | ZNF577, zinc finger protein 577 | VCaP, VCaP-Met | Long read, Short read, qRT-PCR |
| *MBTPS2-YY2** | Class V: Read-through | Xp22.1-p22.2 | MBTPS2, membrane-bound transcription factor peptidase, site 2 | Xp22.2-p22.1 | YY2, YY2 transcription factor | VCaP, LNCaP, VCaP-Met | Long read, Short read, qRT-PCR |
| *C19ORF25-APC2* | Class V: Read-through | 19p13.3 | C19ORF25, chromosome 19 open reading frame 25 | 19p13.3 | APC2, adenomatosis polyposis coli 2 | LNCaP | Long read, Short read, qRT-PCR |
| *WDR55-DND1* | Class V: Read-through | 5q31.3 | WDR55, WD repeat domain 55 | 5q31.3 | DND1, dead end homolog 1 (zebrafish) | RWPE | Long read, Short read, qRT-PCR |
| *SLC45A3-ELK4** | Class V: Read-through | 1q32.1 | SLC45A3, Solute carrier family 45 member 3 | 1q32.1 | ELK4, ETS domain-containing protein | *Met-4* | Short read, qRT-PCR |

# Read-through

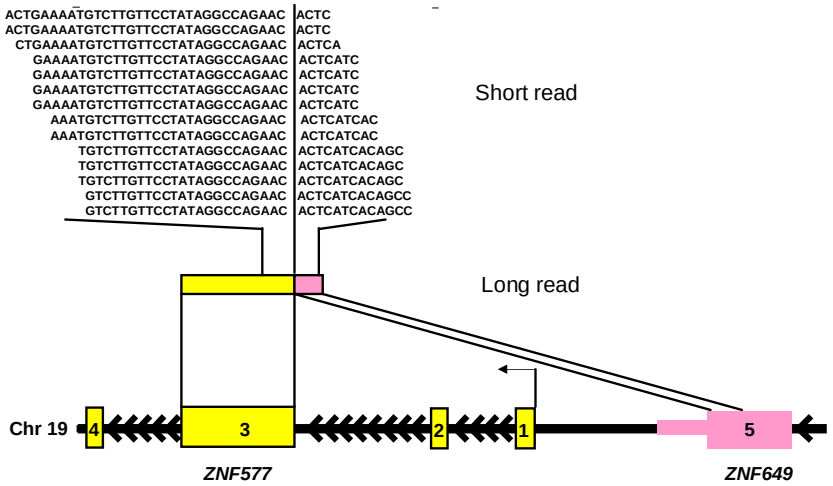
The product of co-transcription of adjacent genes coupled with intergenic splicing (CoTIS)

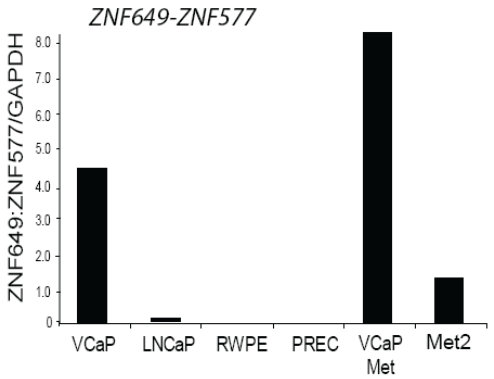



| | Gene fusions |
|---|---|
| VCaP | 1 |
| LNCaP | 3 |
| RWPE | 1 |

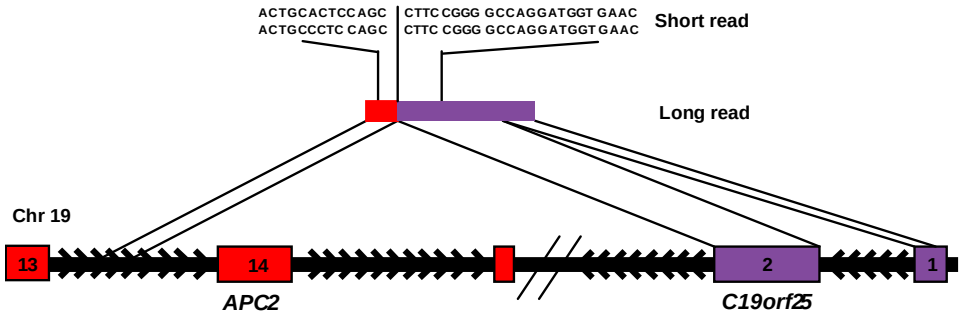
# Read-through Chimeric transcripts

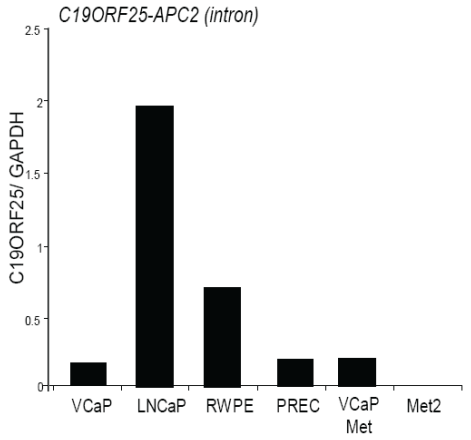
**VCaP**



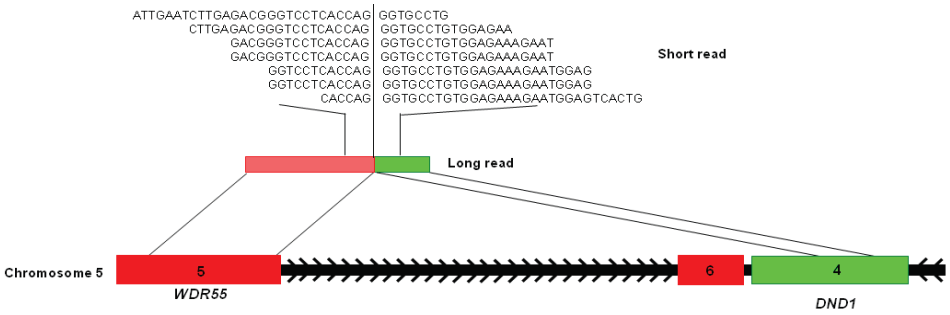



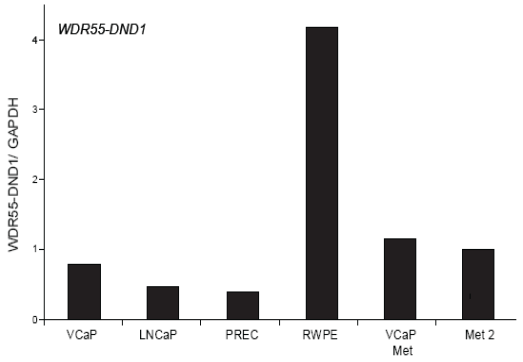

**LNCaP**




**RWPE**

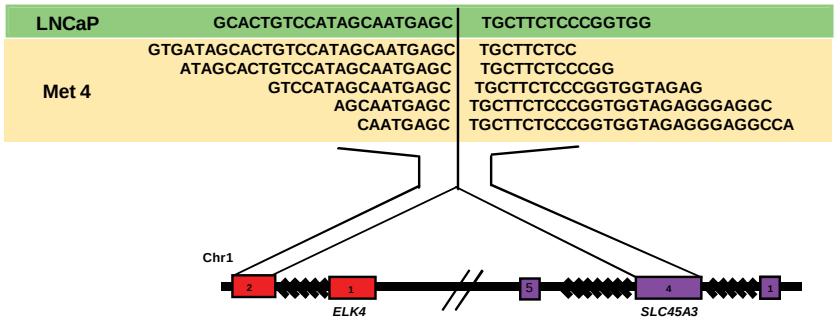


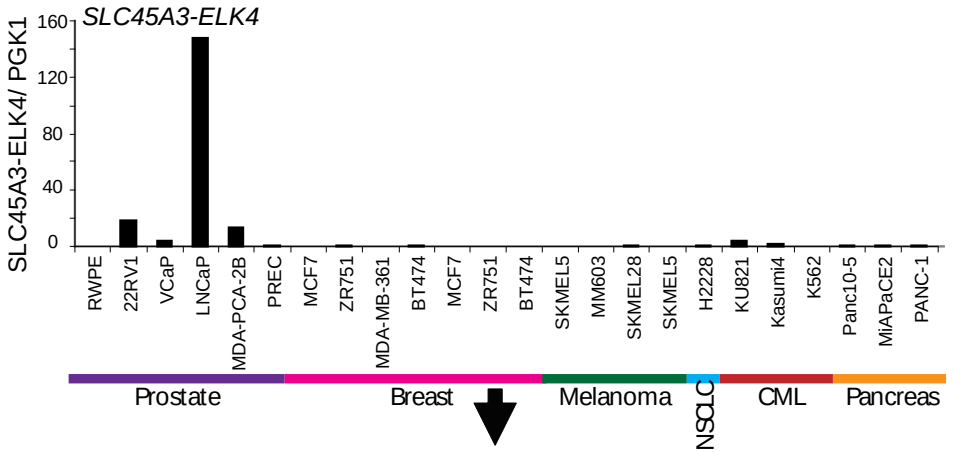



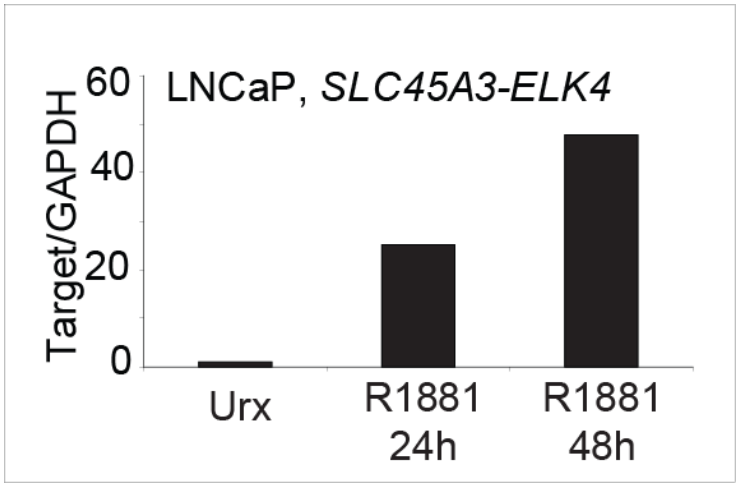
# SLC45A3-ELK4: a Novel, Recurrent, Prostate Specific, Androgen sensitive Chimera

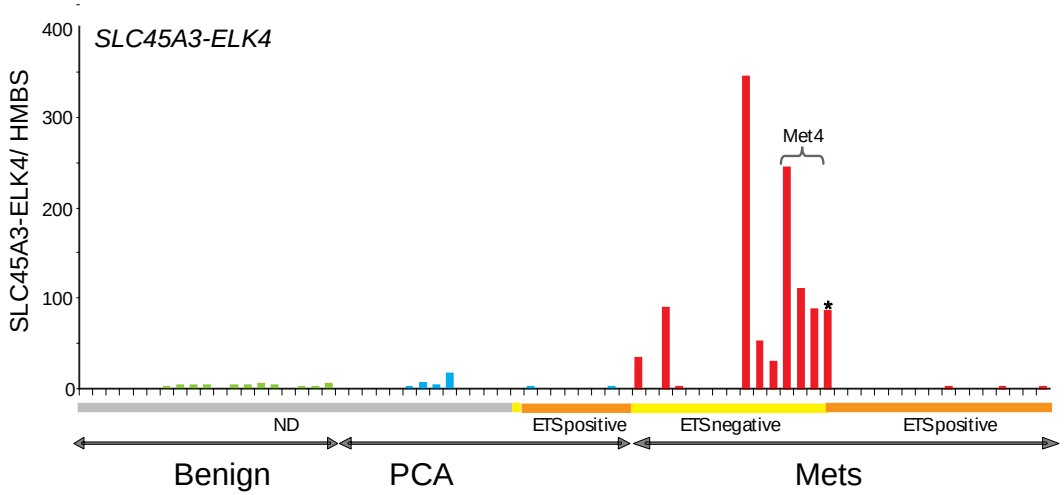
solute carrier family 45, member 3
(Prostate cancer associated protein, Prostein)

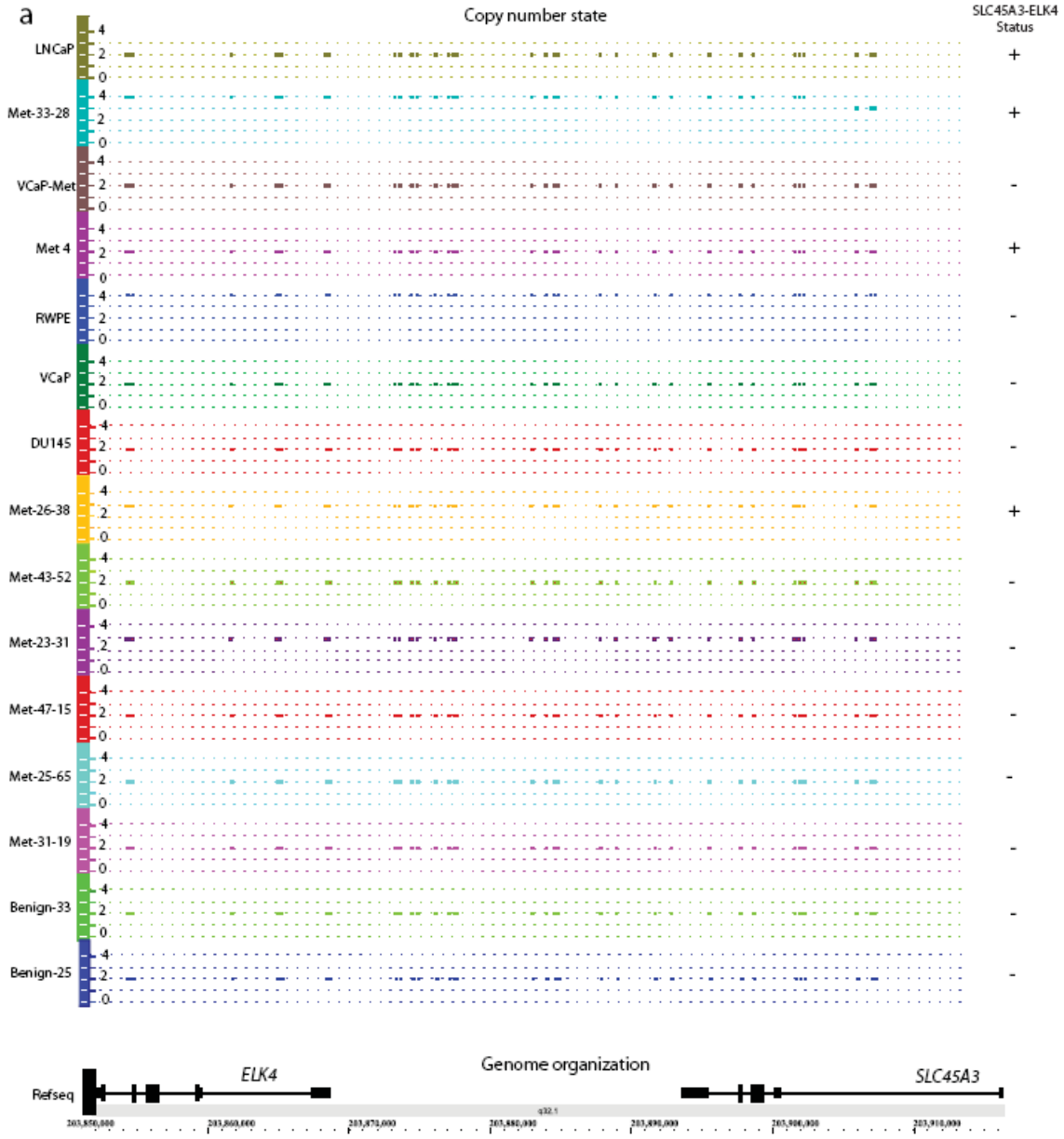
- ELK4, ETS-domain protein
- (SRF accessory protein 1)



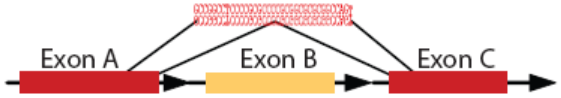






# Lack of microdeletions at the SLC45A3 and ELK4 loci by 1M Affy 6.0 SNP Chips

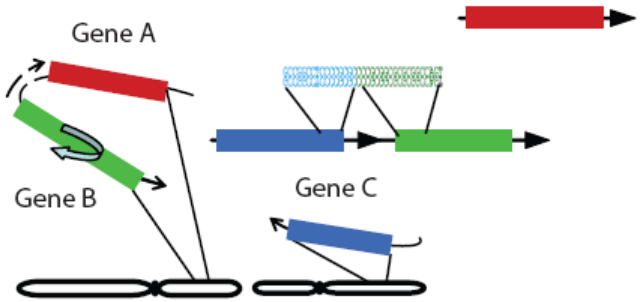
# Chimera classification system

**Alternative Splicing**

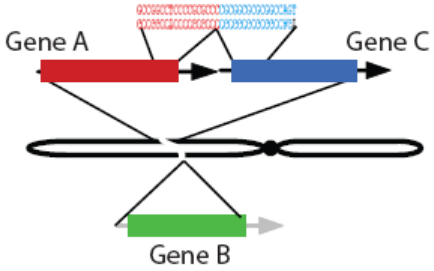



**Class I: Inter-Chromosomal Translocation**

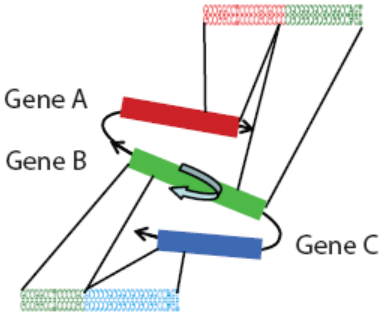

**Class II: Inter-Chromosomal Complex Rearrangements**



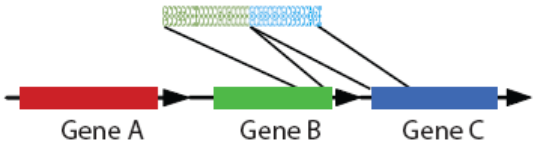

**Class III: Intra-Chromosomal Deletion**



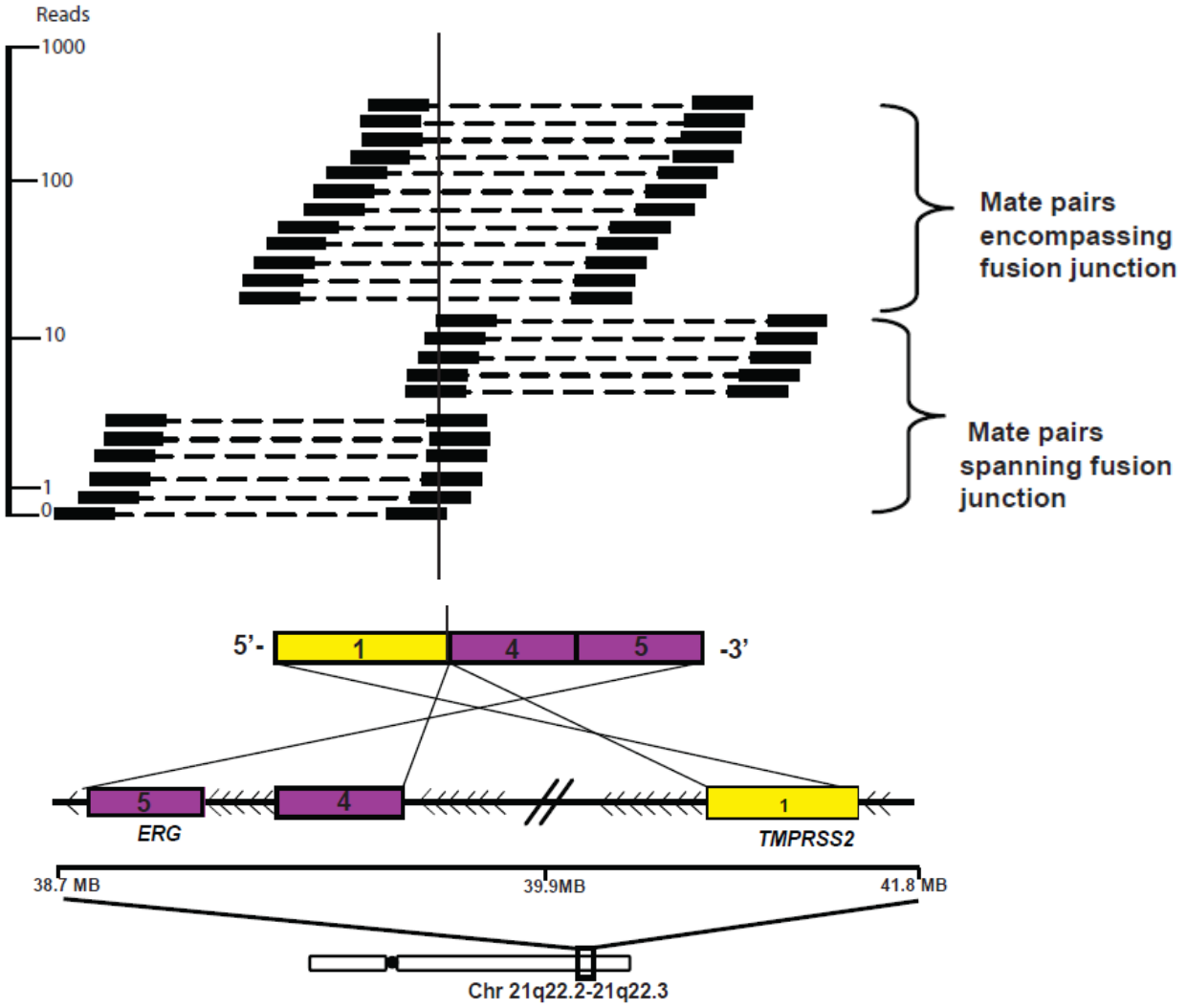

**Class IV: Intra-Chromosomal Complex Rearrangements**



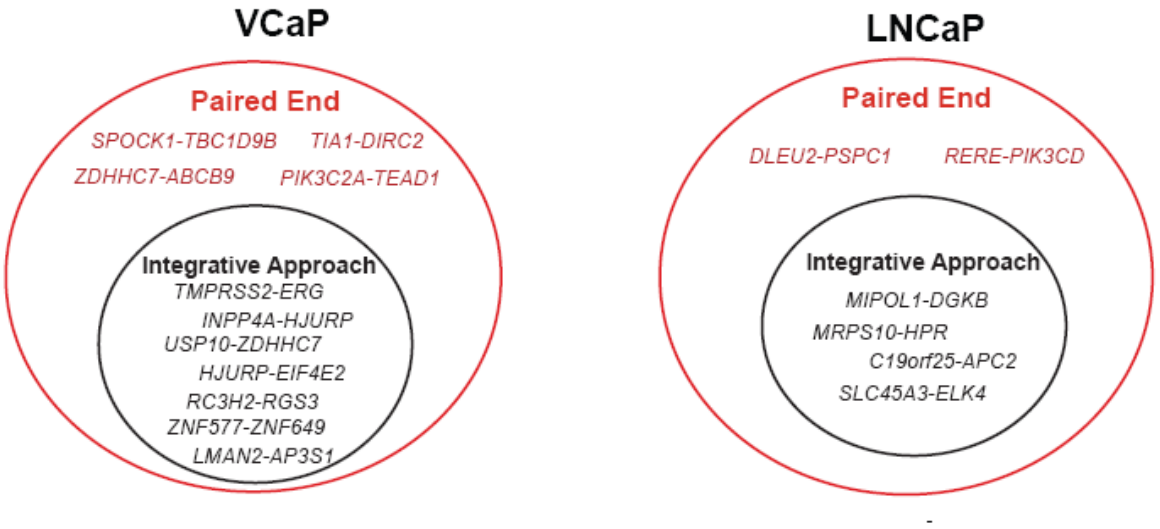

**Class V: Read-throughs**

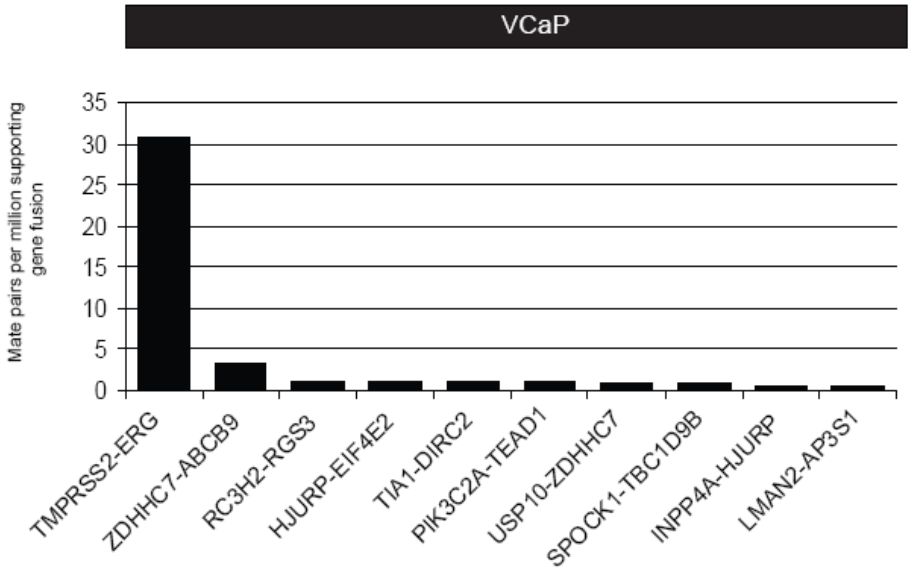


# Paired-end transcriptome strategy for chimera discovery

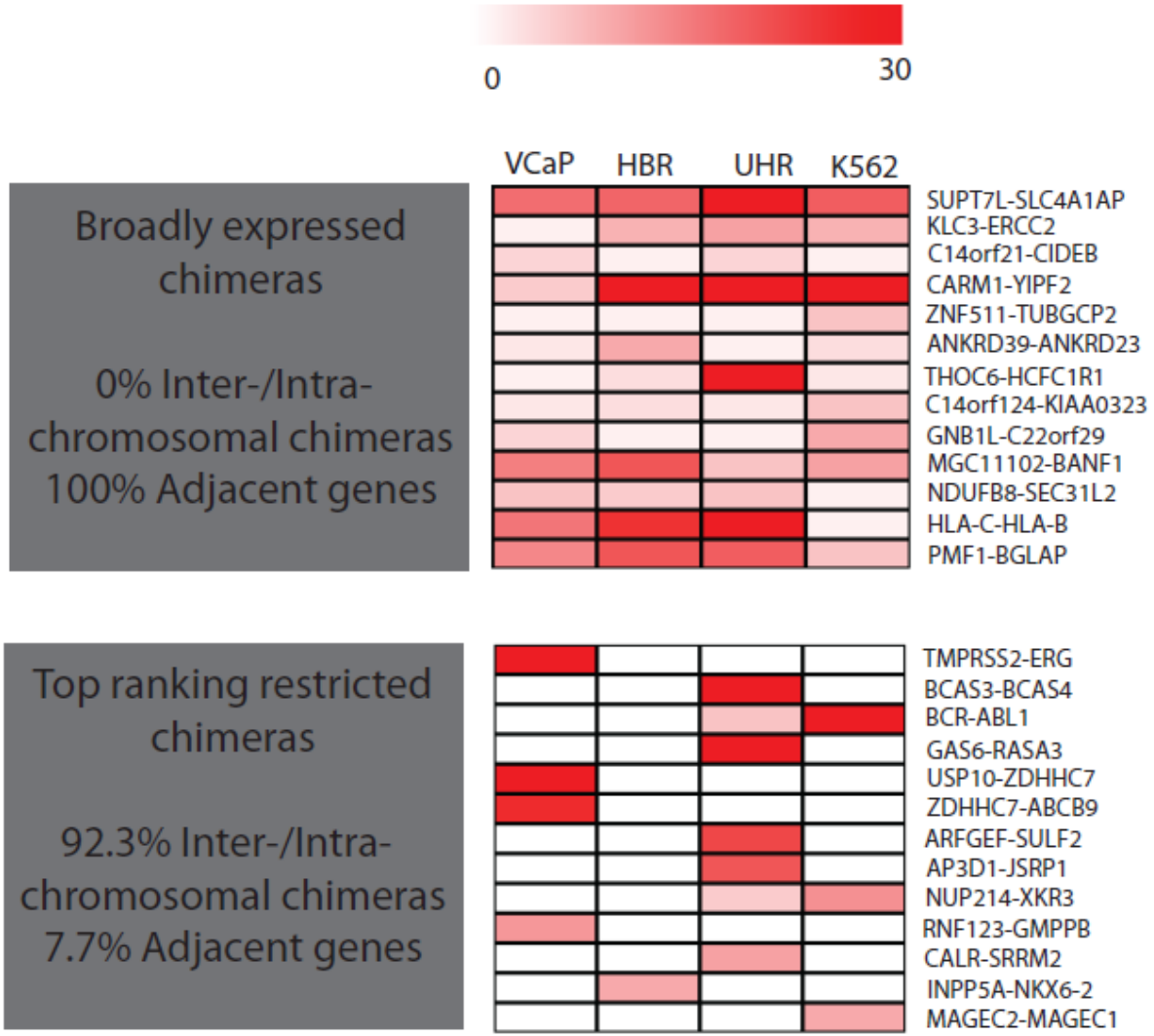


# Comprehensiveness of paired-end sequencing

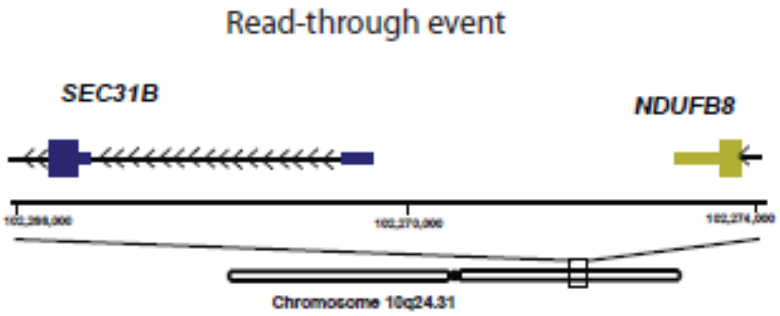


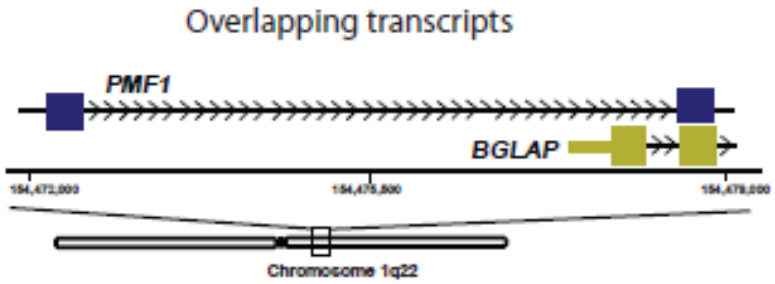



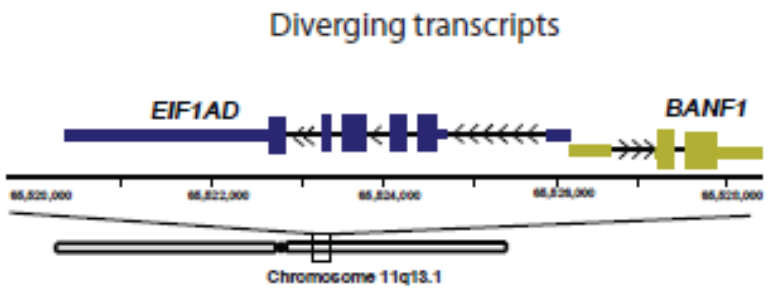
# Broadly Expressed Vs. Restricted Chimeras

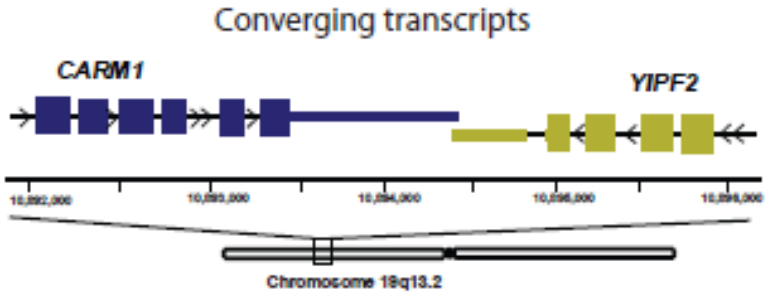


# Additional Classes of Chimeras Easily Observed Using Paired-End Sequencing



# 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

# Gene Fusions in Prostate Cancer


**Scott Tomlins**
**Daniel Rhodes**
**Xuhong Cao**
**Soory Varambally**
**Bharathi Laxman**
**Mohan Dhanasekaran**
**Rohit Mehra**
**Qi Cao**
**Rajal Shah**
**Ken Pienta**
**Jim Montie**
**John Wei**
**Debashis Ghosh**
**Bo Han**
**Anajana Menon**
**Lei Wang**
**Xiaojun Jing**
**Beth Helgeson**
**Jindan Yu**
**Mark Rubin (Cornell)**


# Transcriptome Sequencing for Gene Fusions


**Christopher A. Maher**
**Chandan Kumar-Sinha**
**Xuhong Cao**
**Shanker Kalyana-Sundaram**
**Bo Han**
**Xiaojun Jing**
**Lee Sam**
**Terrence Barrette**
**Nallasivam Palanisamy**

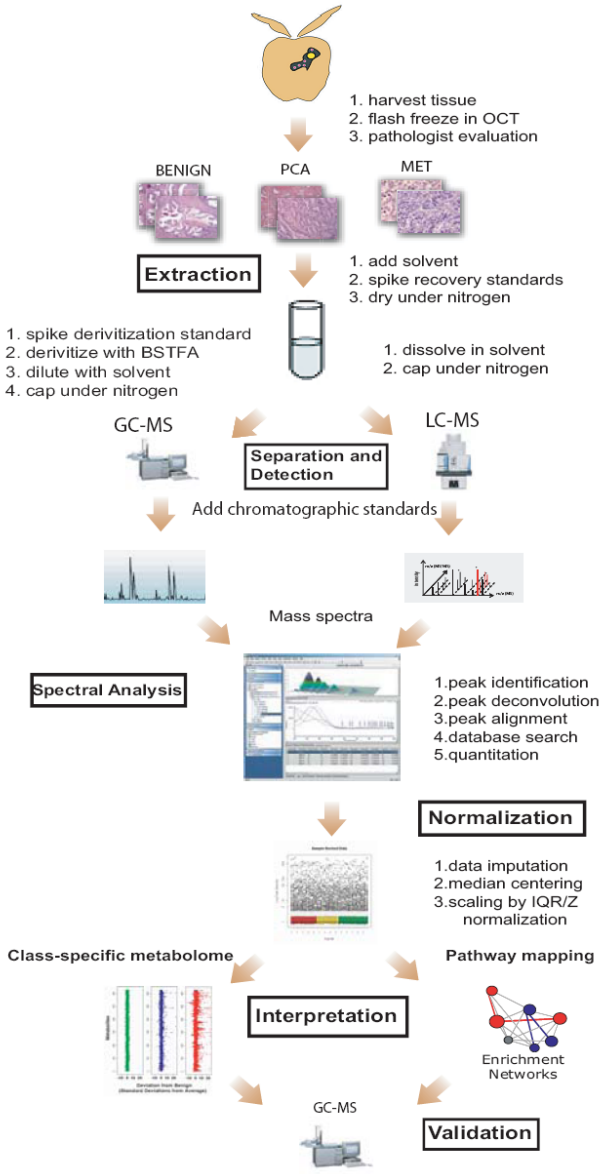


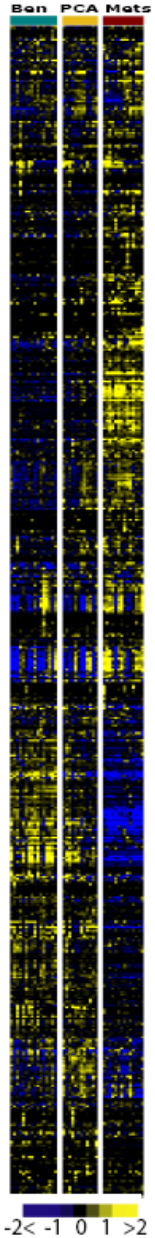
**Funding:**
**HHMI**

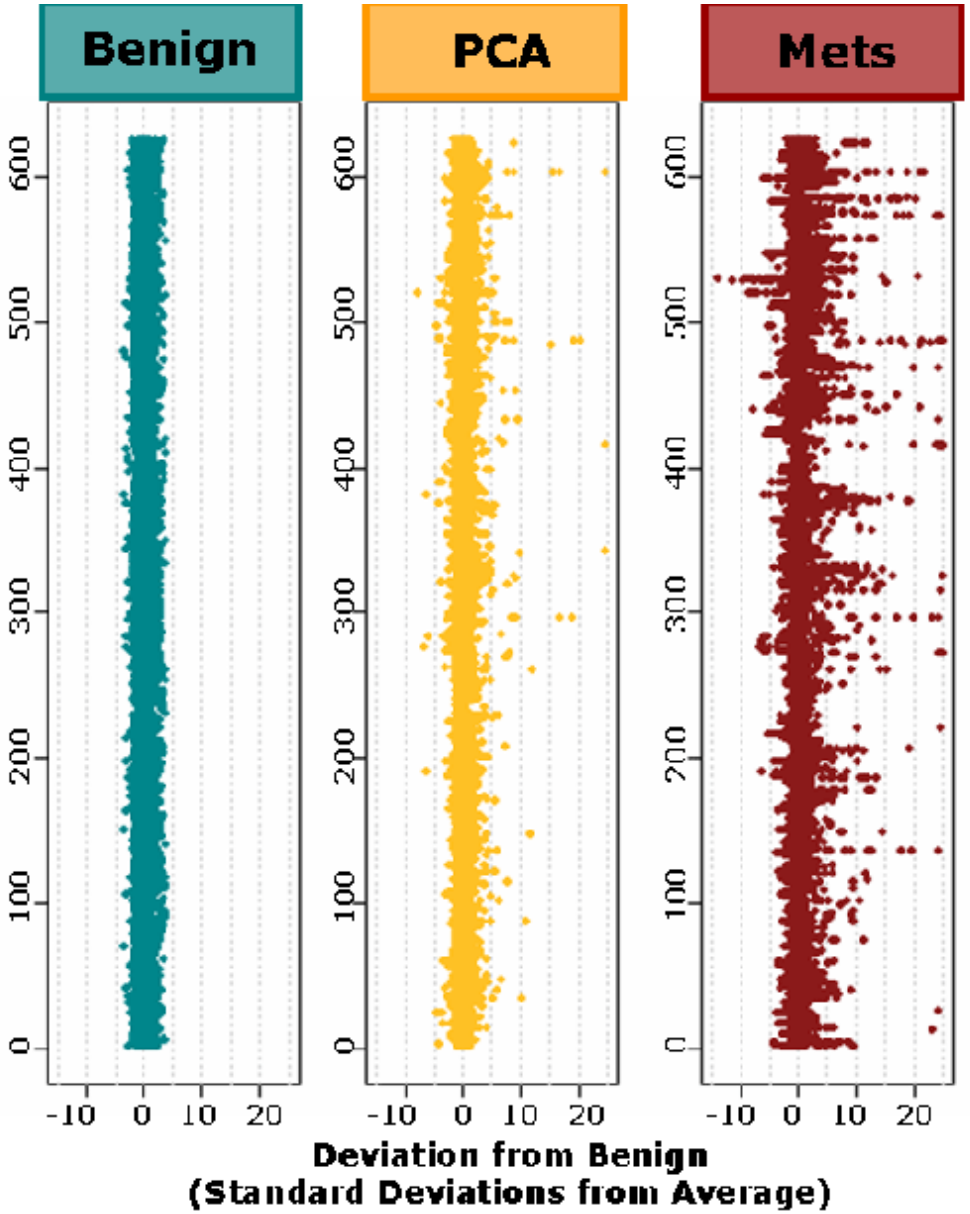


University of Michigan
Center for Translational Pathology

# Metabolomic Profiling of Cancer Progression




Sreekumar et al, *Nature* 2009

# 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

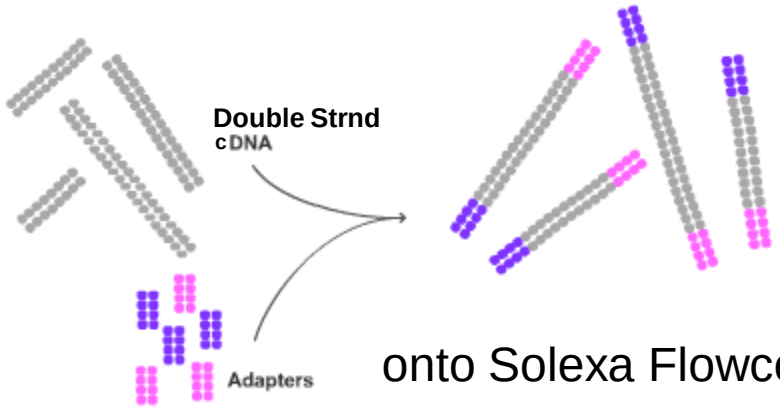
University of Michigan
Medical School

# Next-gen Transcriptome Sequencing

1. mRNA Purification


2. mRNA Fragmentation

3. cDNA Sequencing Library




onto Solexa Flowcell

4. Transcriptome Sequence

**Illumina sequence summary statistics**

| Sample | K562 | | VCaP | | LNCaP | | RWPE | | VCaP-Met | | Met 3 | | Met 4 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Total reads (millions) | 66.9 | | 76.4 | | 57.3 | | 71.9 | | 14 | | 35 | | 9.24 | |
| Pass filter (millions)* | 38.3 | 57.25% | 40.3 | 52.75% | 35.3 | 61.61% | 44.8 | 62.31% | 9.6 | 68.57% | 16.4 | 46.86% | 5.51 | 59.64% |
| Non-mapping reads (millions)** | 2.08 | 5.43% | 12.69 | 31.49% | 1.59 | 4.50% | 1.77 | 3.95% | 0.4 | 4.17% | 1.1 | 6.71% | 0.31 | 5.63% |
| Redundantly mapping reads (millions)** | 1.42 | 3.71% | 1.08 | 2.68% | 1.23 | 3.48% | 1.74 | 3.88% | 0.71 | 7.40% | 1.32 | 8.05% | 0.45 | 8.17% |
| Best hit uniquely maps (millions)** | 19.86 | 51.85% | 15.48 | 38.41% | 19.34 | 54.79% | 26.13 | 58.33% | 7.36 | 76.67% | 12.59 | 76.77% | 4.3 | 78.04% |
| Mitochondrial reads (millions)** | 1.89 | 4.93% | 1.72 | 4.27% | 3.19 | 9.04% | 2.8 | 6.25% | 0.81 | 8.44% | 0.81 | 4.94% | 0.37 | 6.72% |
| Ribosomal reads (millions)** | 13.09 | 34.18% | 9.35 | 23.20% | 10 | 28.33% | 12.34 | 27.54% | 0.31 | 3.23% | 0.62 | 3.78% | 0.09 | 1.64% |

* 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 to due viral sequences
- ~1% of each cell line are categorized as chimera candidates