TRANSLATING SCIENCE TO BETTER HEALTH:
THE POWER OF DIVERSITY AND MULTICULTURAL ENGAGEMENT

CONCURRENT SCIENTIFIC SESSION 9
Molecular and Genetic Markers that Predict Cancer Risk and Clinical Outcomes

DONG LIANG
Texas Southern University
Genetic variants and risk of lung cancer in never smokers: a genome-wide association study

Yuanqing Ye, PhD
Assistant Professor
Department of Epidemiology
Lung Cancer Mortality Rates by Gender and Ethnicity

These rates are based on patients who died in 2005-2009 from 18 SEER geographic areas.

Known Risk Factors of Lung Cancer

Alberg & Samet, Epidemiology of Lung Cancer, Chest, 2005

5-Year Relative Survival Rates

Lung Cancer in Never Smokers (LCINS)

- Never smokers: <100 cigarettes lifetime
- 25% of lung cancer cases are never smoker worldwide
- 15% of men and 53% of women with lung cancer are never smokers worldwide
- ~25,000 LCINS diagnosed in US each year

Death Related to LCINS

Common causes of cancer death in the United States, 2008


LCINS by Gender and Geography Area

Distinct Histology of LCINS

![Bar graph showing proportions of cases for different histologies: Adenocarcinoma (62%), Squamous cell carcinoma (19%), Other (20%).]

Sun et al., Nature Reviews Cancer 2007

Distinct Molecular Features of LCINS

![Bar graph showing proportions of molecular features: KRAS codon 12, 13 mutations (0%), EGFR tyrosine kinase mutation (10%), Total G → T or G → C transitions (20%), P53: G → T transversions (30%), EML4–ALK fusion gene (40%), MGMT methylation (50%).]

Lee et al., Lung Cancer 72 (2011) 9–15

Distinct Survival of LCINS

![Survival curve with details: Non-Smokers: 70% at 60 Mon; Smokers: 10% at 60 Mon.]

Nordquist, Chest 2004

All Rights Reserved - No forms of duplication nor distribution allowed without author's consent
Risk Factors for LCINS

- Older age
- Environmental tobacco smoke (ETS)
- Pre-existing lung disease – Chronic Obstructive Pulmonary Disease (COPD), pneumonia
- Occupational and environmental exposures to carcinogens – Radon exposure and Asbestos
- Oncogenic viruses – human papillomavirus (HPV)

Inherited Genetic Susceptibility of LCINS

- Schwartz et al 1996 observed a 7.2-fold increased risk of developing LCINS for subjects with a positive family history of lung cancer and in the age group of 40 to 59 years
- Yang et al 1999 suggested the presence of high-risk gene contributing to early-onset LCINS
- Wu et al 2004 showed an significantly increased risk for LCINS for subjects with a family history of lung cancer in a study involving 216 Taiwanese female never smokers (OR, 5.7; 95% CI, 1.9 to 16.9)

National Lung Screening Trial (NLST)

- NLST focused on individuals at high risk of LC
  - 55 to 74 years of age
  - Cigarette smoking history of at least 30 pack-years
  - Former smokers must have quit within the past 15 years
- NLST showed a significant reduction in LC mortality of 20%
- Early detection is essential
- Published lung cancer risk prediction models only have modest prediction power
- There is a need to identify novel biomarkers that can provide accurate prediction of individual's risk of lung cancer
### Biomarkers

**Genotype vs. Phenotype Assays**

<table>
<thead>
<tr>
<th>Genotype Assay</th>
<th>Purpose of Assay</th>
<th>Phenotype Assay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Determine the genetic code</td>
<td>Measure gene expression, cellular activity, etc.</td>
<td>Assessing carcinogen metabolism, DNA damage/repair, or enzymatic activity in cultured lymphocytes</td>
</tr>
<tr>
<td>Genotype low penetrance genes (e.g. MMP1, EPHX, CYPT1A1, etc.)</td>
<td>Only require small amounts of DNA; easier to obtain DNA; technically simpler</td>
<td>Live cell lines (e.g. primary blood cultures or EBV-transformed) can be difficult to obtain and costly to maintain</td>
</tr>
<tr>
<td>Assess inherited susceptibility in the genetic code that alter protein expression, function, or localization</td>
<td></td>
<td>Frequently reveal meaningful effects from a constellation of genes (some of which may not be known)</td>
</tr>
</tbody>
</table>

### Single Nucleotide Polymorphisms (SNPs)

- Polymorphisms at which the alleles differ at a single nucleotide in a specific position in the DNA. Usually, there are only two alleles.
- 90% of all human DNA polymorphisms are SNPs.
- Two types of nucleotide base substitutions resulting in SNPs:
  - Transition: substitution between purines (A, G) or between pyrimidines (C, T). Constitute two thirds of all SNPs.
  - Transversion: substitution between a purine and a pyrimidine.

### Genetic Susceptibility to Cancer

- All common cancers exhibit some degree of familial clustering → genetic basis for disease
- Low Penetrance (NAT2, GSTM1, CASP8, FGFR2)
  - Individual Risk Population Attributable Risk
- High Penetrance (BRCA1, BRCA2, MSH2, MLH1)
  - Individual Risk Population Attributable Risk

**MAF:**
- High-frequency SNPs (50%)
- Low-frequency SNPs (5%)
- Rare variants (0.5%)
Common Disease-Common Variant Hypothesis

- Cancer is a complex disease involving multiple genes that have common, low penetrance polymorphisms, as compared with rarer variants but having higher penetrance
- These polymorphisms interact with each other and/or environmental factors to influence disease risk
- “Easy” to screen for the variants in a population

Human Genome Project

- Create genetic and physical maps
- Identify all genes (20,000-25,000)
- Sequence entire genome (3 billion bp)
- Store all information in public databases for free access

The Human HapMap Project

- 2002 International HapMap created
- 2005 Phase 1 completed (~1.3 million SNPs; CEU, YRI, CHB/JPT)
- 2007 Phase 2 published (~3.1 million SNPs; CEU, YRI, CHB/JPT)
- 2009 Phase 3 released (~15.7 million SNPs; 11 populations)
1,000 Genome Project

- Launched in January 2008
- The goal is to characterize over 95% of variants that are in genomic regions accessible to current high-throughput sequencing technologies and that have allele frequency of 1% or higher
- Five major population groups (Europe or European ancestry, East Asia, South Asia, West Africa and the Americas)
- Most recent release contained more than 99M SNPs, INDELs, and SVs

http://www.1000genomes.org

Population Structure

- Utah residents with European ancestry
- Han Chinese in Beijing, China
- Japanese in Tokyo, Japan
- Gujarati Indians in Houston, Texas
- Mexican ancestry in Los Angeles, California

Illumina BeadStudio for Genotyping

<table>
<thead>
<tr>
<th>Illumina chip</th>
<th># of SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>HumanHap550V1</td>
<td>517,383</td>
</tr>
<tr>
<td>HumanHap550V2</td>
<td>518,237</td>
</tr>
<tr>
<td>Human660W</td>
<td>1,072,820</td>
</tr>
<tr>
<td>Human1M</td>
<td>1,999,187</td>
</tr>
<tr>
<td>Human2.5M</td>
<td>2,443,179</td>
</tr>
<tr>
<td>Human5M</td>
<td>4,301,332</td>
</tr>
</tbody>
</table>
Evolution of Genetic Analyses

Methodology

- Hypothesis-driving
- Not dependent on current knowledge
- May identify completely novel markers
- May uncover gene-gene interactions

- Candidate Gene Approach
  - Limited by current knowledge
  - May miss the true variants
  - Low predictive power
  - Limited interaction consideration

- Pathway-based Approach
  - Limited by current knowledge
  - May miss the true variants
  - Low predictive power
  - Limited interaction consideration

- Genome-wide Approach
  - Higher predictive power
  - Consider interaction effects

Candidate Gene Approach

- Hypothesis-driven
- Results easy to interpret
- Small sample size
- Simple analysis
- No multiple comparison issue
- Low cost

Pathway-based Approach

- Hypothesis-driven
- Results easy to interpret
- Small sample size
- Simple analysis
- No multiple comparison issue
- Low cost

GWAS Publications

Published Genome-Wide Associations through 09/2011
1,617 published GWAs at p ≤ 5X10^-8 for 249 traits

GWAS of LCINS

The 5p15.33 Locus Is Associated with Risk of Lung Adenocarcinoma in Never-Smoking Females in Asia

Steps in GWAs

- Genotyping using Whole genome genotyping chip
- Data quality control
- Assessment of population Stratification
- Statistical analysis for association of individual SNPs
- Validation of the identified association through internal/external replication

Study Design

Mayo study
- 377 cases and 545 controls
- 235 18 SNPs

MDACC study
- 328 cases
- 36 residing genes

Harvard study
- 92 cases
- Nearby genes of the top SNPs

UCLA study
- 328 cases
- 407 controls
- 36 residing genes

Analysis of gene expression profiling in never smokers
- Cis-eQTL analysis
- 143 samples from smokers and never smokers

GPC5 expression correlated with rs2352028 and rs2352029
- Gene expression analysis

Study Population

- Mayo Clinic Lung Cancer Study
  - Cases were patients with lung cancer who were classified as never smokers and were identified and recruited between January, 1997, and September, 2008
  - Controls were community residents who were never smokers and were healthy individuals without prior history of cancer (except nonmelanoma skin cancer)
- MD Anderson Lung Cancer Study
  - Cases were patients with newly diagnosed, histologically confirmed primary lung cancer and were recruited from MD Anderson Cancer Center
  - Controls were healthy individuals recruited from the Kelsey-Seybold Clinic, the largest multispecialty physician group practice in the greater Houston area
Study Population

• Harvard Lung Cancer Study
  - Cases were patients with newly diagnosed, histologically confirmed primary lung cancer
  - Controls were healthy non-blood related family members and friends of patients with cancer or patients with cardiothoracic conditions undergoing surgery

• UCLA Lung Cancer Study
  - Cases were patients with histologically confirmed lung cancer, and were identified through the Los Angeles County Cancer Registry administered by the cancer surveillance program at the UCLA
  - The population controls (without history of lung or upper aerodigestive tract cancers) were selected by use of an algorithm to identify eligible controls from a census of each case's neighborhood

Genotyping

• Mayo Clinic Study
  - Illumina HumanHap 370k and 610k BeadChips

• MD Anderson Study
  - Illumina HumanHap 660k and 610k BeadChips

• Harvard Study
  - Illumina Human610-Quad BeadChip

• UCLA study
  - Taqman genotyping

Gene expression profiling Analysis

• Microarray analysis was conducted using Illumina Human Whole Genome DASL array

• After quality control and normalization, a total of 143 pairs of normal and tumor samples from never smoker patients were analyzed

• Of the 143 patients, 70 overlapped with the subjects participating in the Mayo Clinic GWAS study
Quality Control (QC) Steps in GWAS

**Sample QC**
- Remove samples with call rate <95%
- Remove samples with reported sex that did not match with X chromosome heterozygosity
- Remove duplicated samples

**SNP QC**
- Remove SNPs with call rate <95%
- Remove SNPs with departure from HWE in controls ($p<1\times10^{-7}$)
- Remove SNPs with minor allele frequencies (MAF) < 0.005

Population Stratification

- Check for distribution of the p-value using quantile-quantile plots of the observed p-value against expected p-value
- Perform principal components analysis by EIGENSTRAT and adjust the top principal components as covariates

Statistical analysis

- Multivariate conditional logistic regression to assess the effect of SNPs and lung-cancer risk at Mayo
- Multivariate unconditional logistic regression to assess the effect of SNPs and lung-cancer risk at MD Anderson, Harvard, and UCLA
- Meta-analysis to derive the summary estimate of ORs and p values for the combined Mayo, MD Anderson, Harvard and UCLA studies
- Linear regression model to assess the correlation between genotypes and gene expression levels
- Paired t-test to identify genes that were expressed differently in tumour samples and samples of adjacent normal tissue
### Characteristics of study populations

<table>
<thead>
<tr>
<th></th>
<th>Mayo Cases</th>
<th>MD Anderson Cases</th>
<th>Harvard Cases</th>
<th>UCLA Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>control</td>
<td>control</td>
<td>control</td>
<td>control</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>364</td>
<td>364</td>
<td>247</td>
<td>285</td>
</tr>
<tr>
<td>Others</td>
<td>13</td>
<td>13</td>
<td>65</td>
<td>154</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>253</td>
<td>253</td>
<td>211</td>
<td>194</td>
</tr>
<tr>
<td>Male</td>
<td>124</td>
<td>124</td>
<td>131</td>
<td>245</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean(SD)</td>
<td>61.8(14)</td>
<td>61.9(14)</td>
<td>61.4(13)</td>
<td>62.9(12)</td>
</tr>
</tbody>
</table>

### Histological types

<table>
<thead>
<tr>
<th></th>
<th>Mayo</th>
<th>MD Anderson</th>
<th>Harvard</th>
<th>UCLA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenocarcinoma</td>
<td>255</td>
<td>245</td>
<td>57</td>
<td>65</td>
</tr>
<tr>
<td>Squamous</td>
<td>17</td>
<td>26</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Small cell</td>
<td>10</td>
<td>9</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>Carcinoid tumor</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Other NSCLC</td>
<td>35</td>
<td>47</td>
<td>21</td>
<td>18</td>
</tr>
</tbody>
</table>

### Stage

<table>
<thead>
<tr>
<th>Stage</th>
<th>Mayo</th>
<th>MD Anderson</th>
<th>Harvard</th>
<th>UCLA</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>145</td>
<td>92</td>
<td>38</td>
<td>-</td>
</tr>
<tr>
<td>II</td>
<td>19</td>
<td>19</td>
<td>22</td>
<td>-</td>
</tr>
<tr>
<td>III&amp;limited</td>
<td>101</td>
<td>79</td>
<td>22</td>
<td>-</td>
</tr>
<tr>
<td>IV&amp;extensive</td>
<td>145</td>
<td>145</td>
<td>30</td>
<td>-</td>
</tr>
</tbody>
</table>

### GWAS Result by Chromosomes
### Association of rs2352028 With Lung Cancer Risk

#### Overall

<table>
<thead>
<tr>
<th></th>
<th>Case/control</th>
<th>MAF</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mayo study</td>
<td>277/277</td>
<td>0.330/0.23</td>
<td>4.30×10⁻⁵</td>
</tr>
<tr>
<td>MD Anderson study</td>
<td>328/407</td>
<td>0.310/0.23</td>
<td>9.70×10⁻⁶</td>
</tr>
<tr>
<td>Harvard study</td>
<td>92/161</td>
<td>0.270/0.30</td>
<td>7.00×10⁻¹</td>
</tr>
<tr>
<td>UCLA study</td>
<td>514/259</td>
<td>0.330/0.31</td>
<td>9.80×10⁻⁴</td>
</tr>
<tr>
<td>Combined</td>
<td>658/784</td>
<td>0.320/0.26</td>
<td>9.60×10⁻³</td>
</tr>
</tbody>
</table>

#### Adenocarcinoma

<table>
<thead>
<tr>
<th></th>
<th>Case/control</th>
<th>MAF</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mayo study</td>
<td>255/255</td>
<td>0.330/0.23</td>
<td>4.80×10⁻⁵</td>
</tr>
<tr>
<td>MD Anderson study</td>
<td>245/407</td>
<td>0.290/0.23</td>
<td>1.70×10⁻¹</td>
</tr>
<tr>
<td>Harvard study</td>
<td>57/161</td>
<td>0.280/0.30</td>
<td>4.50×10⁻¹</td>
</tr>
<tr>
<td>UCLA study</td>
<td>55/259</td>
<td>0.330/0.31</td>
<td>3.27×10⁻¹</td>
</tr>
<tr>
<td>Combined</td>
<td>477/782</td>
<td>0.310/0.27</td>
<td>3.00×10⁻³</td>
</tr>
</tbody>
</table>

### The 13q31 Region Encompassing rs2352028

- Presented at the 13th RCMI International Symposium on Health Disparities | December 9-13, 2012 | San Juan, Puerto Rico
- All Rights Reserved - No forms of duplication nor distribution allowed without author's consent
Expression analysis of GPC5 by genotype

Conclusion

- Genetic variants at 13q31.3 are associated with susceptibility to lung cancer in never smokers
- These variants also alter the expression of GPC5
- The transcription level of GPC5 in normal lung tissue was twice that detected in matched lung adenocarcinoma tissue
- Down regulation of GPC5 might contribute to the development of lung cancer in never smokers

Acknowledgement

MD Anderson Cancer Center
  - Xifeng Wu
  - Jie Lin
  - Margaret R Spitz
UCLA
  - Zuo-Feng Zhang
  - Shen-Chih Chang
  - Renyi Wang
Harvard University
  - David C Christiani
  - Chau-Chyun Sheu
  - Li Su
  - Feng Chen
Mayo Clinic
  - Ping Yang
  - Yafei Li
  - Mariza de Andrade
  - Liang Wang
  - Marie C Aubry
  - Jeremiah A Aakre
  - Mark S Allen
  - Julie M Cunningham
  - Jin Jen
NCI
  - Curtis C Harris
Molecular epidemiology studies of breast cancer in Puerto Rico

Jaime L. Matta, Ph.D.
Professor
Department of Pharmacology, Physiology and Toxicology
Ponce School of Medicine and Health Sciences

ACKNOWLEDGEMENTS

This study is being supported by grants from the NCI Center to Reduce Health Disparities and NIH-MBRS Program grant S06 GM008239-20 to Ponce School of Medicine through Dr. J. Matta.

Co-Investigators
-- Manuel Bayona, MD, MS, PhD.
-- Carolina Alvarez, MD, MS, PhD.

Collaborators
-- Julie Dutil, PhD
-- Eric Suarez, PhD
-- Rafael Guerrero, Dr.Ph.
-- David Sidransky, MD

Oncologists, Surgeons, Plastic Surgeons and Gynecologists through the island
-- Miguel Echenique, MD - San Juan
-- Aníbal Torres, MD – Ponce, Yauco
-- Guillermo Bolaños, MD - Ponce
-- Felipe Sánchez Gaetan, MD - Ponce
-- Eduardo Ramirez Lizardi, MD - Ponce
-- José Ortiz Rosado, MD - Ponce
-- Juan González Cruz, MD - Ponce
-- Jesús Monasterio, MD - Ponce
-- José Cangiano, MD - Ponce
-- Ángel Romero Sánchez, MD - Ponce
-- Francisco García Rivera, MD - Ponce

Laboratory Staff
-- Hannia Delgado, B.S.S.
-- Luisa Morales, B.S.
-- Wanda Vargas, RN

Ph.D. Students
-- Carmen Ortiz, B.S.

Dr.PH Students
-- Yeidyly Vergne, B.S.
Specific Aims

• Specific Aim 1: To test whether DNA repair capacity (DRC) of lymphocytes can be utilized as a predictor for breast cancer risk in women of different age groups with special focus on invasive ductal breast carcinoma.

• Specific Aim 2: To identify whether the expression of DNA repair genes that are associated with various phenotypes of DNA repair capacity vary as a function of age.

• Specific Aim 3: To study key epidemiological risk factors for breast cancer and their association with DNA repair capacity in women.

• Specific Aim 4: To develop scientific expertise in the area of epigenetics in order to utilize these tools in ongoing cancer studies.
Results

Hypothesis: A low DNA repair is a significant risk factor for breast cancer.

The association of DNA Repair with breast cancer risk in women. A comparative observational study
<table>
<thead>
<tr>
<th>Variable</th>
<th>Cases</th>
<th>Controls</th>
<th>OR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRC1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>245</td>
<td>81</td>
<td>43.9 (24.4, 79.1)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Medium</td>
<td>94</td>
<td>233</td>
<td>5.1 (2.9, 9.1)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>High</td>
<td>39</td>
<td>286</td>
<td>Referent</td>
<td></td>
</tr>
<tr>
<td>Smoke</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>50</td>
<td>54</td>
<td>2.0 (1.1, 3.9)</td>
<td>0.034</td>
</tr>
<tr>
<td>No</td>
<td>334</td>
<td>544</td>
<td>Referent</td>
<td></td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>186</td>
<td>402</td>
<td>Referent</td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>81</td>
<td>105</td>
<td>1.6 (1.0, 2.6)</td>
<td>0.057</td>
</tr>
<tr>
<td>Divorced</td>
<td>48</td>
<td>73</td>
<td>1.1 (0.6, 1.9)</td>
<td>0.770</td>
</tr>
<tr>
<td>Widow</td>
<td>356</td>
<td>19</td>
<td>4.9 (2.1, 11.6)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 – 8</td>
<td>24</td>
<td>9</td>
<td>5.5 (1.8, 16.2)</td>
<td>0.002</td>
</tr>
<tr>
<td>9 – 12</td>
<td>130</td>
<td>159</td>
<td>1.5 (1.0, 2.3)</td>
<td>0.048</td>
</tr>
<tr>
<td>Associate or higher degree</td>
<td>170</td>
<td>408</td>
<td>Referent</td>
<td></td>
</tr>
<tr>
<td>Family history of cancer (not BC)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>255</td>
<td>354</td>
<td>1.4 (1.0, 2.0)</td>
<td>0.047</td>
</tr>
<tr>
<td>No</td>
<td>130</td>
<td>252</td>
<td>Referent</td>
<td></td>
</tr>
<tr>
<td>BC history in any family member</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>102</td>
<td>95</td>
<td>2.0 (1.2, 3.2)</td>
<td>0.010</td>
</tr>
<tr>
<td>No</td>
<td>283</td>
<td>511</td>
<td>Referent</td>
<td></td>
</tr>
<tr>
<td>Consumed vitamins last 5 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>218</td>
<td>404</td>
<td>0.7 (0.5, 1.0)</td>
<td>0.049</td>
</tr>
<tr>
<td>No</td>
<td>164</td>
<td>197</td>
<td>Referent</td>
<td></td>
</tr>
<tr>
<td>Taking vitamins now</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>202</td>
<td>382</td>
<td>0.7 (0.4, 0.9)</td>
<td>0.041</td>
</tr>
<tr>
<td>No</td>
<td>181</td>
<td>215</td>
<td>Referent</td>
<td></td>
</tr>
<tr>
<td>Calcium</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>77</td>
<td>167</td>
<td>0.7 (0.4, 1.0)</td>
<td>0.076</td>
</tr>
<tr>
<td>No</td>
<td>304</td>
<td>431</td>
<td>Referent</td>
<td></td>
</tr>
</tbody>
</table>

* Adjusted by age, BMI, family history of breast cancer, menopause, number of children, alcohol use, smoking, multivitamin use, marital status, saturated fat consumption and DRC.

Which DNA repair genes in breast tumors are associated with dysregulation of DNA repair capacity?
DNA repair candidate gene expression in breast cancer tumors

23 DNA Repair genes significantly expressed, p < 0.05

Receptors in breast cancer and potential associations with DNA Repair Capacity

Figure 1: Association of Breast Cancer Subtypes and Selected Variables with Low Distributions in DNA Repair genes. Adjusted OR and 95% CI.
Summary

- Women with BC have a DRC 60% lower that women without breast cancer.
- Selected risks factors for BC showed an association with DRC.
- 23 candidate DNA repair genes were found to be differentially expressed in the 35 breast tumors analyzed \( (p<0.001) \).
- Triple negative (ER-, PR-, Her2neu-) receptor status were associated with a low DNA repair capacity. Efficient DRC is essential to minimize cell processes such as mutations, replication errors, and genomic instability. Therefore, low DRC might play a role in the development of receptor-specific tumors.

Future directions, questions, challenges

1. Whether DRC deficiency in patients with breast cancer preexisting, acquired during oncogenesis, or some combination of the two is still to be determined.
2. Need to conduct prospective studies to determine:
   A. whether women free of breast cancer with a low DRC have a higher incidence of breast cancer as compared to those with high DRC,
   B. and whether women with breast cancer and a low DRC are at an increased risk of having a recurrent disease as compared to those with high DRC.
3. Development of a new assay to measure DNA Repair Capacity (major challenge) is an ongoing process.
4. Study epigenetic mechanisms associated with the regulation of DRC. Latimer et al. 2010 found that the majority of NER genes are down-regulated in early-stage breast cancer; however, whether such epigenetic down-regulation is secondary to mutation at a known NER structural gene or an unknown NER regulatory gene needs to be determined.
5. Vision of the role of DNA Repair as integral part of personalized medicine not only in cancer, aging, but in all diseases.
BRCA2

BRCA2 or the breast cancer type 2 susceptibility protein in humans is a 3418 amino acid protein localized in the nucleus.

BRCA2 is a tumor suppressor gene.

BRCA2 is ubiquitously expressed in all eukaryotes except few exceptions like yeast.

Role of BRCA2 has been implicated in the development of breast, ovarian, prostate, pancreatic and esophageal cancer.

Evidence is accumulating that silencing or dysfunction of BRCA2 gene, might also be important in the pathogenesis of a significant proportion of sporadic, non-familial breast cancers.

BRCA2 is involved in maintenance of genome stability, specifically the homologous recombination pathway for double-strand DNA repair.
Double-stranded break (DSB) of DNA caused by
- DNA replication
- DNA repair
- Normal environment
- Programmed recombination

Often associated with mutation, predisposition to cancer or cell death

BRCA2/Rad51
Non-Homologous Recombination
Homologous Recombination

Normal recovery
Normal cells
No mutation

Role of BRCA2 in DSB repair by Homologous Recombination (HR)

BRCA2 expression is silenced at the quiescent stage and desilenced when cells enter the replicative phase.

The mechanisms of cell cycle-dependent regulation of BRCA2 gene expression is not well understood.

We are working on two distinct mechanisms of BRCA2 gene regulation in the human breast cancer cells:

1. SLUG-dependent mechanism operative in SLUG-high breast cancer cells.
2. SLUG-independent mechanism operative in the SLUG-negative breast cancer cells.
Regulation of BRCA2 gene expression by the novel transcription factor ZAR2

Genomic context of human BRCA2 and ZAR2 gene promoters.

Through the GeneRacer technique we found that the bi-directional promoter of BRCA2 and ZAR2 gene are overlapping type. We also identified a new exon 1 for ZAR2 gene.

Amino acid sequence and domain map of human ZAR2 protein. RBD: RNA binding domain; ZF: Zinc finger.
Luciferase assay for the evaluation of BRCA2 forward and reverse promoter activities in synchronized human breast cancer cells

Human BRCA2 gene promoter forward and reverse activities are differentially regulated during the cell cycle

Real-time RT-PCR evaluation of the relative levels of BRCA2 and ZAR2 mRNAs in different human breast cancer cells at G0/G1 and S/G2 phases.

Immunofluorescence microscopy using FLAG antibody detecting FLAG-tagged ZAR2 protein in the synchronized MCF7 cells.
These observations suggest that not only the expression of ZAR2 gene is strictly regulated cell cycle-dependently but its subcellular localization is also controlled in a growth stage-dependent manner.

In order to understand the role of cell cycle-dependent regulation of ZAR2 gene expression on BRCA2 levels, we explored whether the C4-type zinc finger protein ZAR2 can bind to the BRCA2 promoter DNA.

In vivo binding of ZAR2 protein to the BRCA2/ZAR2 gene promoter by ChiP using FLAG antibody binding to the C-terminal FLAG-tagged ZAR2 protein.

ZAR2 binds to BRCA2/ZAR2 gene promoter in vivo preferentially at the G0/G1 phase.
**Regulation of BRCA2 Gene Expression**

**A. SLUG-positive breast cancer cells**
- SLUG is translocated to the nucleus in the non-dividing cells.
- SLUG binds to E2 box sequence in the BRCA2 silencer region (-791 bp to -821 bp).
- SLUG binds to CtBP1 and recruits HDAC1 to heterochromatinize the BRCA2 gene promoter.
- Silencing of BRCA2 gene expression in non-dividing cells.

**B. SLUG negative breast cancer cells**
- ZAR2 is transcribed by the reverse activity of BRCA2 promoter (-187 bp to 310 bp).
- ZAR2 is translocated to the nucleus in the non-dividing cells.
- ZAR2 in the nucleus binds to the dsRNA (111 bp) generated by the overlapping ZAR2 and BRCA2 mRNAs.
- Guided by the RNA, ZAR2/RNA complex binds to the BRCA2/ZAR2 bidirectional promoter.
- ZAR2 then recruits DNMT1 which recruits MBD1 and HDAC1 to heterochromatinize the BRCA2 gene promoter.
- Silencing of BRCA2 gene expression in non-dividing cells.

**Summary**

BRCA2 gene promoter has bi-directional activity, expressing BRCA2 and a novel C4-type zinc finger containing transcription factor ZAR2.

Subcellular location of ZAR2 and its expression from the reverse promoter of the BRCA2 gene are stringently regulated in a cell cycle dependent manner.

ZAR2 binds to BRCA2/ZAR2 bidirectional promoter in vivo and is responsible, at least in part, for the silencing of BRCA2 gene expression in the G0/G1 phase in human breast cells.

ZAR2 is over expressed in the tumor tissues compared to the adjacent normal tissues.

**Acknowledgments**

Prof. Gautam Chaudhuri
Dr. Mukul Mittal

Meharry Medical College Morphology Core: Dr. S. Goodwin

Funding from NIH 5U54 RR026140-03(NCRR)/ 8U54 MD007593-03 (NIMHD) to SM and DOD-CDMRP IDEA Grant# BC050641 to GC
Osteosarcomas are highly aggressive malignant neoplasms:

At the time of diagnosis:
1. Most are already classified as high grade
2. Are poorly differentiated
3. Approximately 20% have detectable metastasis, mostly pulmonary

Prognosis is very poor: only about 10% of osteosarcoma patients achieve long-term (5 years) disease-free intervals
What do osteosarcomas, retinoblastomas, and small cell lung carcinomas have in common?

1. Highly aggressive

2. Non-functional adherens junctions
   - Retinoblastomas: detachment of the cadherin/catenin complex from the cortical cytoskeleton (Van Aken et al., 2002)
   - Osteosarcoma: weak, cytoplasmic localization of adherens junction proteins (Kashima et al., 1999)
   - Small cell lung carcinomas: Cytoplasmic β-catenin localization (Rodríguez-Salas et al., 2002)

3. High frequency of mutations in the RB1 locus
   - Codes for the retinoblastoma tumor suppressor protein (pRb)

pRb as a cell regulator of cell proliferation and cell adhesion

Bone-specific RB1 knock-out mice

- RB+/+
- RB-/-

EZRIN
GAPDH

All Rights Reserved - No forms of duplication nor distribution allowed without author's consent
RB1 loss results in abnormal cadherin expression

RB1 loss results in global deregulation of cell adhesion related genes

 Canonical Pathways Deregulated by RB1 Deficiency. MetaCore cellular pathway analysis in conjunction with the Canonical Pathway database from GeneGo, Inc., was used to determine the top 10 cellular processes affected by pRb deficiency and the associated p-value. The final column of the table represents the total number of objects present in the canonical pathway that were affected by pRb deficiency and the total number of affected genes.
Molecular mechanisms linking Rb to cell adhesion

Rb re-expression in SAOS-2 cells

Fold-change induced by Rb (log10)

Fold Activation by Rb (log 10 scale)

Transcriptional regulation of PAK1 expression by Rb
Disruption of cell adhesion due to Rb loss may be linked to the molecular etiology of lung cancer

Exploration of the Director’s Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma Database revealed that several Rb-activated cell adhesion genes (as identified by our microarrays) are correlated with increased overall survival.

Conclusions

1. Rb controls cell adhesion by at least two mechanisms:
   - Transcriptional control of cell adhesion genes (e.g., cadherins)
   - Transcriptional repression of PAK1, possibly by blocking E2F’s transactivation by E2F

2. Loss of cell adhesion engendered by Rb loss may be related to the molecular etiology of human lung adenocarcinoma (work in progress).

Acknowledgements

Santiago Lab:
Bernadette Sosa and Viviana Vázquez (graduate students),
Joan Fred (Lab technician)

Collaborators:
Dr. Philip W. Hinds (Tufts Medical Center, Boston, MA)
Dr. W. Douglas Cress (Moffitt Cancer Center (Tampla, FL))

Funding:
Ponce School of Medicine and Health Sciences start-up funds (RCMI Grant G12 RR003050)
ACS Institutional Research Grant 93-032-13
NIH NCI U54CA118809, and NIH NCI 3 U54 CA118809-04-S1 (ARRA supplement)
- NIH NCI 1U54CA163071-01A1