Family Studies in Shanxi, China: Search for high-risk susceptibility genes

Alisa M. Goldstein
Genetic Epidemiology Branch/DCEG
NCI-IARC Tumor Workshop: ESCC
September 2016
## Risk Variant Classification

<table>
<thead>
<tr>
<th>Variant Type</th>
<th>Frequency</th>
<th>Penetrance Risk*</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-risk</td>
<td>Very rare to Rare</td>
<td>Very High [5 - &gt;20]</td>
</tr>
<tr>
<td><strong>Family Studies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate-risk</td>
<td>Rare [often &lt;0.01]</td>
<td>Intermediate [1.5 - &lt;5]</td>
</tr>
<tr>
<td><strong>Family Studies/GWAS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low-risk</td>
<td>Common</td>
<td>Low [&lt;1.5]</td>
</tr>
<tr>
<td><strong>GWAS</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Relative risk of carriers versus noncarriers
Areas in red are those with highest esophageal cancer rates from 1973-1975 national mortality survey.
Family History (FH) of ESCC in First-Degree Relatives (FDR) of 600 ESCC Cases/1514 Controls

[Gao et al, 2009 BMC Cancer]

<table>
<thead>
<tr>
<th>Relatives</th>
<th>Ctrl/Case%</th>
<th>OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>FH of ESCC (FDR)</td>
<td>6/16%</td>
<td>2.84</td>
<td>2.09-3.86</td>
</tr>
<tr>
<td>Father</td>
<td>3/7%</td>
<td>2.01</td>
<td>1.31-3.10</td>
</tr>
<tr>
<td>Mother</td>
<td>2/6%</td>
<td>3.27</td>
<td>1.96-5.47</td>
</tr>
<tr>
<td>Siblings</td>
<td>1/6%</td>
<td>4.66</td>
<td>2.67-8.13</td>
</tr>
<tr>
<td>1 relative with ESCC</td>
<td>6/14%</td>
<td>2.53</td>
<td>1.84-3.50</td>
</tr>
<tr>
<td>2+ relatives with ESCC</td>
<td>0.3/2%</td>
<td>10.0</td>
<td>3.24-31.2</td>
</tr>
<tr>
<td>Non-blood relatives</td>
<td>0.6/1%</td>
<td>1.86</td>
<td>0.68-5.10</td>
</tr>
</tbody>
</table>
### Family History (FH) of Cancer in First-Degree Relatives of Cases/Controls*

[**Gao et al, 2009 BMC Cancer**]

<table>
<thead>
<tr>
<th>FH of Cancer</th>
<th>ESCC OR (95% CI)</th>
<th>GCA OR (95% CI)</th>
<th>GNCA OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any UGI Cancer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Affected</td>
<td>2.1 (1.6-2.7)</td>
<td>1.4 (1.1-1.9)</td>
<td>1.5 (1.1-2.2)</td>
</tr>
<tr>
<td>≥2 Affected</td>
<td>6.4 (2.8-15.6)</td>
<td>5.4 (2.3-12.7)</td>
<td>4.8 (1.7-13.6)</td>
</tr>
<tr>
<td>Any non-UGI Cancer</td>
<td>1.0 (0.7-1.3)</td>
<td>0.9 (0.7-1.3)</td>
<td>1.2 (0.8-1.8)</td>
</tr>
</tbody>
</table>

*600 ESCC ca, 598 GCA ca, 316 GNCA ca, 1514 controls
Shanxi Family Study

- Number of families: 148
  - Number of UGI cases/family: 2 - >8
    - Most families have 1-2 UGI cases with DNA

- 92 families with ≥3 UGI cases selected for whole exome sequencing (WES)
  - 24 families: ESCC only
  - 53 families: ESCC & Gastric cancer (GC)
    - 177 UGI cases/obligate carriers
      - 75% families ≤2 UGI cases with DNA
Tylosis with Esophageal Cancer

- Genetic disorder characterized by thickening of the palms and soles (hyperkeratosis), white patches in the mouth (oral leukoplakia), and very high risk of ESCC
- Disease gene: *RHBDF2* (17q25)

Is tylosis seen in Shanxi? No

- No disease-related mutations in *RHBDF2*
UGI WES Analysis Strategies

Although multiple candidate genes in 1-2 families, no frequent high-risk susceptibility genes identified

- **Exclude common variants seen in public databases or in-house datasets**
- **Apply dominant genetic model**
- **Same variants/genes in multiple families**
  - *In Silico* function prediction/literature review
- **Technical Validation/Co-segregation** 5-100

**Follow-up in Shanxi case-control sample**
- Sequenced 21 genes
- ~560 Cases
- ~590 Controls

**Follow-up in Henan case-control sample**
- Genotyped ~50 variants
- ~2600 ESCC cases
- ~2000 GC cases
- ~4500 Controls
Next Steps

- Determination of top candidate genes/variants for further follow-up

- **Follow-up activities:**
  - Functional studies of top genes/variants
    - In silico
    - Laboratory-based
  - Genotyping/sequencing of more cases and/or controls
  - Collaborations to share WES UGI data
    - Other populations with ESCC families?
High-Risk Gene Discovery Challenges

- Genetic heterogeneity – multiple genes responsible for disease susceptibility
- Lack of power/informativeness of sample
- Underlying complexity of disease cause(s)
  - Clinical/epidemiologic heterogeneity and/or misclassification
- Rarity of mutations
  - Proving causality for “private” mutations
  - Differentiating true positives from false positives
- No (or few) high-risk susceptibility genes
Genetic Risk Loci

- **High**
  - Rare alleles causing Mendelian disease

- **Intermediate**
  - Low-frequency variants with intermediate effect

- **Modest**
  - Rare variants of small effect very hard to identify by genetic means

- **Low**
  - Very rare

- **Common**
  - Few examples of high-effect common variants influencing common disease
  - Common variants implicated in common disease by GWA

Slide courtesy of Teri Manolio, NHGRI
Recent evaluation in complex diseases suggested that multiple low-risk variants may explain disease risk in some families

- Examine in exome families using OmniXpress

Polygenic risk score (PRS) in ESCC

- Low-risk ESCC loci identified from GWAS
- Tested whether family history (FH) subgroups differed in extent of enrichment for polygenic effects
  - Shanxi ESCC FH+ vs FH- cases
    - Average PRS: 0.55 vs 0.49 (p=0.04)
To search for high-risk genes in ESCC, we conducted a family study in Shanxi, China

WES performed in 92 3+ UGI case families
- Including 24 ESCC only and 53 ESCC + GC

No frequent high-risk genes yet identified
- Multiple candidate genes in 1-2 families each
  - Extensive follow-up including functional studies required to prove causation

Examination of low-risk variants in WES families in process to determine whether risks in families result from multiple low-risk variants
Difficult Questions

- Although WES offers opportunities for identifying high-risk ESCC genes, studies are complex with many challenges
  - How to prove causality for “private” mutations?
- What are the best strategies for identifying high-risk ESCC genes?
- Do multiple low-risk variants explain disease risk in some families from high-risk regions?
- How do we reduce risk/prevent disease in ESCC families from high-risk regions?
Acknowledgements

- DCEG, NCI
  - Philip R. Taylor
  - Nan Hu
  - Wen-Qing Li
  - Hua Su
  - Lemin Wang
  - Chaoyu Wang
  - Melissa Rotunno
  - Jianxin Shi
  - Paula Hyland
  - Hyuna Sung
  - Margaret A. Tucker
  - Stephen J. Chanock
  - Geoffrey S. Tobias

- Shanxi, China
  - Ding Ti
  - He Li-Ji,
  - Han Xaio-Yu

- CGR, NCI
  - Cancer Genomics
  - Research Laboratory,
  - Leidos Biomedical
  - Research, Inc.

- Henan, China
  - Li-Dong Wang
  - Xin Song

- IMS
  - Carol Giffen

- CCR, NCI
  - Maxwell Lee