Insights: Identification of Candidate Variants Using Exome Data in Ophthalmic Genetics

3/7/2013
Khanh-Nhat Tran-Viet, MHA
Agenda

• Landscape of NGS
• Overview of our research
• Using Exome data to filter-
  • Step by step application of
techniques (Tips and Tricks )
    • Filter by Marker Statistics
    • Filter by Gene List
    • Filter by Functional Predictions
    • Filter by Multiple Columns
• Challenges/Opportunities
Terri Young, MD, MBA
Professor of Ophthalmology, Pediatrics and Medicine
-Duke University Center for Human Genetics (CHG)
-Duke Eye Center
-Duke-National University of Singapore -Graduate Medical School
Our research interests

- Myopia (nearsightedness)
- Primary Congenital Glaucoma
- Stickler/Wagner Syndromes
- Corneal Dystrophy
- Strabismus (crossed eye)
- Eyelid Malformation
- Microphthalmia/Anopthalmia
Why Vision Research?

Children
- The most common vision seen in children are:
  - Myopia (near-sightedness)
  - Hyperopia (far-sightedness)
  - Astigmatism (unequal curves of the eye)
  - Anisometropia (unequal eyes)

Causes
- Genetic
- Twin Studies
- Animal Studies
- Environmental
- Epidemiology Studies

Economic Impact
- Glare from contact lenses, glasses, etc. → $200 annually in US
- Myopia-related costs exceed $14 billion dollars annually
Economic Impact

Graph 1.1  Total Annual Economic Impact of Vision Problems in the U.S.

Total: $51.4 billion

<table>
<thead>
<tr>
<th>Costs (in billions)</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct medical costs ($16.2)</td>
<td></td>
</tr>
<tr>
<td>Other direct costs ($11.2)</td>
<td></td>
</tr>
<tr>
<td>Lost productivity ($8.0)</td>
<td></td>
</tr>
<tr>
<td>Medical care expenditures ($5.12)</td>
<td></td>
</tr>
<tr>
<td>Informal care costs ($0.36)</td>
<td></td>
</tr>
<tr>
<td>Health utility costs ($10.5)</td>
<td></td>
</tr>
</tbody>
</table>

Doctor visits, contacts, glasses, etc --> $200 annually in US

Myopia related costs exceeds $14 billion dollars annually
Children

The most common types seen in children are:

Myopia (nearsightedness)

Strabismus (crossed eyes)

Amblyopia (lazy eye)
Prevalence of myopia

- 19.45% - India (Dandona et al. 2002)
- 21.8% - China (Xu et al. 2005)
- 33.1% - US (Vitale et al. 2008)
- 84% - Taiwan (Lin et al. 2004)
- 85% - Hong Kong (Lam et al. 2004)

- 3% - Prevalence of high myopia (≤-6D) (Vongphanit et al. 2002)
Causes

- Genetic
  - Twin Studies
  - Animal studies

- Environmental
  - Epidemiology studies
Stickler Syndrome

Clinically variable and heterogeneous disorder
Prevalence of 1 in 10,000
- Ocular
- Auditory
- Skeletal
- Orofacial

Family
- 49 members (16 affected)
- Initially diagnosed Wagner Syndrome
- Variable clinical presentations
- Screened for known mutations

Genes
- COL2A1
- COL11A1
- USH2A
- COL1A1
- COL1A2

Processes

Input
- 2 affected in Family
- NGS performed
- Variants detected

Bioinformatics
- Aligned, removed duplicates
- GATK to call variants
- BCF variants

Output
- NC-variant Coding Rate for analysis
- BCF variants for visualization

Filtering Assumptions
- Rare
- Autosomal Dominant model
- Coding-
  - Nonsynonymous
  - Conserved
  - Biologically relevant
Stickler Syndrome
Clinically variable and heterogeneous disorder

Prevalence of 1 in 10,000

- Ocular
- Auditory
- Skeletal
- Orofacial
# Genes

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>% of Disease Attributed to Mutations in This Gene</th>
<th>Inheritance</th>
</tr>
</thead>
<tbody>
<tr>
<td>COL2A1</td>
<td>80%-90%</td>
<td>Autosomal Dominant</td>
</tr>
<tr>
<td>COL11A1</td>
<td>10%-20%</td>
<td>Autosomal Dominant</td>
</tr>
<tr>
<td>COL11A2</td>
<td>Rare, unknown</td>
<td>Autosomal Dominant</td>
</tr>
<tr>
<td>COL9A1</td>
<td>Rare, unknown</td>
<td>Autosomal Recessive</td>
</tr>
<tr>
<td>COL9A2</td>
<td>Rare, unknown</td>
<td>Autosomal Recessive</td>
</tr>
</tbody>
</table>

*GeneReviews*
Family

- 49 members (16 affecteds)
- Initially diagnosed Wagner Syndrome
- Variable clinical presentations
- Screened for known mutations

<table>
<thead>
<tr>
<th>Clinical manifestations</th>
<th>Seen in at least 1 clinically affected family member</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cataract</td>
<td>Yes</td>
</tr>
<tr>
<td>Nuclear sclerosis</td>
<td>Yes</td>
</tr>
<tr>
<td>Avascular sheets</td>
<td>Yes</td>
</tr>
<tr>
<td>Optically empty vitreous</td>
<td>Yes</td>
</tr>
<tr>
<td>Horsehoe retinal tear</td>
<td>Yes</td>
</tr>
<tr>
<td>Retinal detachment</td>
<td>Yes</td>
</tr>
<tr>
<td>Chorioretinal scars</td>
<td>Yes</td>
</tr>
<tr>
<td>Myopia</td>
<td>Yes</td>
</tr>
<tr>
<td>Scoliosis</td>
<td>Yes</td>
</tr>
<tr>
<td>Flexible Joint hypermobility</td>
<td>Yes</td>
</tr>
<tr>
<td>Sensorineural or conductive hearing loss</td>
<td>No</td>
</tr>
<tr>
<td>Cleft palate</td>
<td>No</td>
</tr>
<tr>
<td>Mild epiphyseal dysplasia</td>
<td>No</td>
</tr>
<tr>
<td>Congenital megalophthalmos</td>
<td>No</td>
</tr>
</tbody>
</table>

- 2 affected males
- Nine affected females
- Cataract
- HiFi
- Humoral (Homo)
Filtering Assumptions

- Rare
- Autosomal Dominant model
- Coding-
  Nonsynonymous
- Conserved
- Biologically relevant
73,360 variants
Filter by annotation track membership Exon +/- 10bp
(Removes majority of intronic, intergenic, etc)

Common dbSNP137
(Filter out variants which have MAF > 1%)

Variant Classification
(Classifies variants – ie synonymous, coding, etc etc)

Filter by Marker Statistics
(Allows to set criteria: which sample should have what genotype)

Filter by Gene List

Filter by Multiple Columns

Filter by 7 Functional Predictions

25,485
3,071
1,900
314
58
9
Filter by Marker Statistics

Use genotype information:
- Call rate
  "X number of my samples have to have some sort of genotype call"
- Frequencies of genotype of interest
  "X samples have/don’t have ___ genotype calls"

Genotype Statistics by Marker

(No variable is set as dependent.)

Classify alleles by allele frequency
Classify alleles by reference/alternate
(Reference field in map: "Reference")

Marker Statistics
- Call rate (fraction not missing)
- Number of Alleles
- Allele frequencies
- Hardy-Weinberg Equilibrium (HWE) P-Value
- Fisher’s exact test for HWE P-Value
- Signed HWE R (positive if more homozygous)

Count Tables
- Genotype counts
- Allele counts

Additional Output
- Output -log10(Value)
Variant Classification
(Classifies variants – ie synonymous, coding, etc etc)

Filter by Marker Statistics
(Allows to set criteria: which sample should have what genotype)

Filter by Gene List

Filter by 7 Functional Predictions

**Filter by Multiple Columns

1,900

314

58

9
Filter by gene list

Use list of candidate genes, associated genes, etc as a tool to filter
Variant Classification
(Classifies variants – i.e. synonymous, coding, etc etc)

Filter by Marker Statistics
(Allows to set criteria: which sample should have what genotype)

Filter by Gene List

Filter by 7 Functional Predictions

**Filter by Multiple Columns

1,900
314
58
9
Filter by NS Functional Prediction

* in silico protein prediction
  - SIFT
  - PolyPhen2
  - MutationTaster
  - Mutation Assessor
  - FATHMM

* in silico conservation
  - GERP++
  - PhyloP

<table>
<thead>
<tr>
<th>Map</th>
<th>Marker</th>
<th>C10</th>
<th>R11</th>
<th>C12</th>
<th>R13</th>
<th>C14</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1:186062678-SNV</td>
<td>0.39</td>
<td>Tolerated</td>
<td>0.969</td>
<td>Probably Damaging</td>
<td>0.989477</td>
</tr>
<tr>
<td>2</td>
<td>1:243493888-SNV</td>
<td>0.32</td>
<td>Tolerated</td>
<td>0.137</td>
<td>Benign</td>
<td>0.000315</td>
</tr>
<tr>
<td>3</td>
<td>2:27601843-SNV</td>
<td>0.32</td>
<td>Tolerated</td>
<td>0.165</td>
<td>Benign</td>
<td>0.053192</td>
</tr>
<tr>
<td>4</td>
<td>2:96954854-SNV</td>
<td>0.07</td>
<td>Tolerated</td>
<td>0.464</td>
<td>Possibly Damaging</td>
<td>0.999981</td>
</tr>
<tr>
<td>5</td>
<td>4:20598044-SNV</td>
<td>0.08</td>
<td>Tolerated</td>
<td>0.8</td>
<td>Possibly Damaging</td>
<td>0.999923</td>
</tr>
<tr>
<td>6</td>
<td>4:37863193-SNV</td>
<td>0.02</td>
<td>Damaging</td>
<td>0.058</td>
<td>Benign</td>
<td>0.976697</td>
</tr>
<tr>
<td>7</td>
<td>4:46314633-SNV</td>
<td>0.09</td>
<td>Tolerated</td>
<td>0.434</td>
<td>Benign</td>
<td>0.407169</td>
</tr>
<tr>
<td>8</td>
<td>4:47556908-SNV</td>
<td>1</td>
<td>Tolerated</td>
<td>0.042</td>
<td>Benign</td>
<td>0.002124</td>
</tr>
</tbody>
</table>
Filter by multiple columns

"at minimum, X number of categories I set must be met"

<table>
<thead>
<tr>
<th>MutationAssessor</th>
<th>MutationAssessor Pred</th>
<th>FATHMM Score</th>
<th>FATHMM Pred</th>
<th>GERP++ RS</th>
<th>PhyloP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.24</td>
<td>Predicted Non-Functional (Neutral)</td>
<td>-1.17</td>
<td>Tolerated</td>
<td>5.04</td>
<td>2.353</td>
</tr>
<tr>
<td>0.345</td>
<td>Predicted Non-Functional (Neutral)</td>
<td>0.87</td>
<td>Tolerated</td>
<td>3.75</td>
<td>1.428</td>
</tr>
<tr>
<td>0.205</td>
<td>Predicted Non-Functional (Neutral)</td>
<td>0.38</td>
<td>Tolerated</td>
<td>3.59</td>
<td>1.099</td>
</tr>
<tr>
<td>3.13</td>
<td>Predicted Functional (Medium)</td>
<td>0.2</td>
<td>Tolerated</td>
<td>5.5</td>
<td>2.308</td>
</tr>
<tr>
<td>0.93</td>
<td>Predicted Non-Functional (Low)</td>
<td>-3.7</td>
<td>Damaging</td>
<td>2.93</td>
<td>0.859</td>
</tr>
<tr>
<td>2.215</td>
<td>Predicted Functional (Medium)</td>
<td>1.94</td>
<td>Tolerated</td>
<td>4.92</td>
<td>2.367</td>
</tr>
<tr>
<td>1.14</td>
<td>Predicted Non-Functional (Low)</td>
<td>-1.05</td>
<td>Tolerated</td>
<td>5.95</td>
<td>2.817</td>
</tr>
<tr>
<td>0.425</td>
<td>Predicted Non-Functional (Neutral)</td>
<td>1.26</td>
<td>Tolerated</td>
<td>4.29</td>
<td>2.565</td>
</tr>
</tbody>
</table>
73,360 variants
Filter by annotation track membership Exon +/- 10bp
(Removes majority of intronic, intergenic, etc)

Common dbSNP137
(Filter out variants which have MAF > 1%)

Variant Classification
(Classifies variants – ie synonymous, coding, etc etc)

Filter by Marker Statistics
(Allows to set criteria: which sample should have what genotype)

Filter by Gene List

Filter by 7 Functional Predictions

**Filter by Multiple Columns**

25,485
3,071
1,900
314
58
9
Results

- A Novel nonsense mutation on exon 2 of COL2A1
- Not present in 2000 chromosomes nor in any public databases, highly conserved
- Exon 2 mutations is predominantly ocular-only phenotype
- Currently in submission
Challenges and Opportunities

Filtering strategies will depend on your phenotype:

1. Not what's right or wrong... but think "efficiency"
2. Filter to minimize re-analysis
3. Understanding the study design
- Filtering strategies will depend on your phenotype
- Not what's right or wrong...but think "efficiency"
- Filter to minimize re-analysis
- Understanding the study design

**Using Golden Helix SVS**

*Pros:*

- Customer service
- User-friendly

*Cons:*

- Learning curve
Filter by annotation track membership Exon +/- 10bp 
(Removes majority of intronic, intergenic, etc)

Common dbSNP137
(Filter out variants which have MAF > 1%)

Variant Classification
(Classifies variants – ie synonymous, coding, etc etc)

Filter by Marker Statistics
(Allows to set criteria: which sample should have what genotype)

Filter by Gene List

Filter by 6 Functional Predictions

Filter by 6 Functional Predictions

Filter by all known dbSNP?

Look for runs of Homozygosity

Filter using QC metrics

Compound Heterozygous analysis

Filter by Recessive model

PCA, association tests
Filtering strategies will depend on your phenotype

- Not what's right or wrong... but think "efficiency"

- Filter to minimize re-analysis

- Understanding the study design

Using Golden Helix SVS
Using Golden Helix SVS

Pros:
- Customer service
- Opportunities for everyone to participate
- Documentation/Internal QC
- Endless possibilities

Cons:
- Learning curve
- Cost prohibitive
- Continual updates
Fast

Good

Cheap
Acknowledgements

Center for Human Genetics
Terri Young, MD, MBA
Erica Nading, CGC
Diana Abbott, PhD
Yi-Ju Li, PhD
Xiaoyan Luo, PhD

Vincent Soler, MD
Caldwell Powell
Elizabeth St.Germain
Kristina Nicholson
Jenny Wei
Krystina Quow

Duke-NUS
Steve Rozen, PhD
John McPherson, PhD
Thomas Klemm
Stuart Tompson, PhD

Golden Helix
Greta Peterson, PhD
Autumn Laughbaum

A*STAR Singapore
Sebastian Maurer-Stroh, PhD
Vachiranee Limviphuvadh, PhD

Collaborators
Max Johnson, MD
Charles Johnson
Benjamin Bakall, MD
Ed Stone, MD
Barathi Veluchamy, PhD
Ravikanth Metlapally, PhD
Tammy Yanovitch, MD

Hudson Alpha
Shawn Levy, PhD
Braden Boone, PhD
Jack Wimbish
Questions?

Contact: Khanh-Nhat Tran-Viet
K.tranviet@duke.edu