Genetics and risk for associated autoimmune disorders and refractory coeliac disease

Alexandra Zhernakova/ Cisca Wijmenga
University Medical Center Groningen
The Netherlands
c.wijmenga@umcg.nl
Celiac disease is a classical complex disorder.
Gluten is not the sole cause of celiac disease

- Other non-HLA genes
- HLA-DQ2/DQ8

Gluten + Other environmental factors

40% of heritability + 60% of heritability

Equal to Celiac Disease

HLA-DQ2/8 necessary, but not sufficient (OR ~ 5)
Genetics of celiac disease has progressed slowly.

Candidate gene association studies

Linkage analysis

<table>
<thead>
<tr>
<th>HLA</th>
<th>CTLA4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1973</td>
<td>1997</td>
</tr>
<tr>
<td>1998</td>
<td>1999</td>
</tr>
<tr>
<td>2000</td>
<td>2001</td>
</tr>
<tr>
<td>2002</td>
<td>2003</td>
</tr>
<tr>
<td>2004</td>
<td>2005</td>
</tr>
<tr>
<td>2006</td>
<td>2007</td>
</tr>
<tr>
<td>2008</td>
<td>2009</td>
</tr>
</tbody>
</table>
Genetics of celiac disease has progressed slowly

Candidate gene association studies

- HLA
- CTLA4

Genome-wide association (GWA) studies

Linkage analysis

Years:
- 1973
- 1997
- 1998
- 1999
- 2000
- 2001
- 2002
- 2003
- 2004
- 2005
- 2006
- 2007
- 2008
- 2009
... in 1000s of patients and controls

100-1000s of patients (cases)

DNA from cases

SNP ‘chip’ (>500,000 SNPs)
One for each case DNA

Compare differences to discover SNPs associated to disease

100s-1000s of healthy individuals (controls)

DNA from controls

SNP ‘chip’ (>500,000 SNPs)
One for each control DNA
GWAS have led to a revolution in human genetics (complex diseases)
GWAS have led to a revolution in human genetics (complex diseases)
GWAS have led to a revolution in human genetics (complex diseases)
GWAS have led to a revolution in human genetics (complex diseases)
GWAS have led to a revolution in human genetics (complex diseases)

June 2011: 1449 loci at $p \leq 5 \times 10^{-8}$ for 237 traits
Celiac disease: one of the GWAS success stories to identify non-HLA risk factors

A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21

Newly identified genetic risk variants for celiac disease related to the immune response

Multiple common variants for celiac disease influence immune gene expression
What did we learn from GWAS?

Looking insight the associated loci...
Celiac disease-associated variants are common and show modest effect (OR<1.5)

Hrdličková et al., submitted
HLA and non-HLA variants can explain ~50% of genetic variation
The 39 non-HLA loci represent 115 genes

Genes with a function in the immune system are depicted in blue (25 loci, 44 genes)
25 celiac disease-associated GWAS loci overlap with other immune-mediated loci.

Gutierrez-Achury et al., J Int Med 2011
Immune genes fall into 5 broad categories:
- Cell signaling
- Cell migration
- Cytokines
- NFκB activity
- T cells

Gutierrez-Achury et al., J Int Med 2011
Can this explain the extra-intestinal manifestations?

Trynka et al., Trends Mol Med 2010
We should be careful. Loci may look the same but could be different ...

Trynka et al., Gut 2009
GWAS is done. What’s next?

• Genetics
  • Narrowing the region
  • Identification of the casual variant(s)
    • Fine-mapping
    • Cross-population studies
    • Cross-disease studies
  • Missing heritability
  • Low frequency variants

• Identification of downstream effects
  • Pathway analysis
  • Expression studies
ImmunoChip – fine-mapping and shared immune genetics

- 12 AID study groups: T1D, IBD, RA, CoeID, MS, SLE...
- 189 distinct loci associated to immune-related diseases
- 1kG (and individual efforts) sequence variants
- Dense HLA genotyping
- All together – 200K samples
10-20x greater SNP density than Hap550 (post-QC)
13 new loci identified, making the total number of non-HLA loci 39

Novel loci = blue
Loci with multiple signals = grey highlight
~50% of cases: Multiple common independent SNPs, including rare variants (< 5%)

13 loci show independent second (or third) signal

54 independent non-HLA signals in 39 loci

29 signals could be localized to a single gene
Cross ethnic mapping

We can narrow down several loci, including IL2-IL21 block
So far one paper was published from Immunochip data

- **In celiac disease!**

---

**Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease**

Cross-disease meta-analysis - co-morbid diseases

- Rheumatoid arthritis
- Multiple sclerosis
- Type 1 diabetes
- Celiac disease
- Crohn’s disease
- Ulcerative Colitis
- Asthma
- Psoriasis
- Ankylosing spondylitis
- SLE
- Autoimmune thyroid disease
Meta-analysis: 14 shared CeD-RA loci, including 4 new loci

Meta-Analysis of Genome-Wide Association Studies in Celiac Disease and Rheumatoid Arthritis Identifies Fourteen Non-HLA Shared Loci

Implicates antigen presentation and T-cell activation as a shared mechanism
Downstream effect – how severe the damage is?

Deleterious mutation? Or regulatory variants?
Most association signals are non-coding.

Only 3 protein-altering variants were identified.
Approx. 50% of associated SNPs show eQTL effects.
The strongest associated SNP from the associated region has a *cis*-regulatory effect on *IL18RAP* expression; The celiac disease risk allele is associated with low transcript levels of this gene.
Example of an eQTL from the celiac region on 2q12.1

The strongest associated SNP from the associated region has a cis-regulatory effect on IL18RAP expression; The celiac disease risk allele is associated with low transcript levels of this gene.
Building the co-expression networks helps establishing the key pathways

Network construction using co-expression of genes
Point to genes with potential rare risk variants?
Prevention from uncontrolled self-reactivity = tolerance

Importance of the thymus
THEMIS, TNFRSF14, RUNX3, ETS1

Innate immune detection of viral RNA
UBE2E3, BACH2, TNFAIP3, TLR7, TLR8

T-and B-cell co-stimulation
CTLA4, ICOS-CD28, TNFRSF14, CD80, ICOSLG, TNFRSF9, TNFSF4

Role for rotavirus infections?

- These processes are partly shared with other immune-mediated diseases
- There are still many more genetic factors to be discovered
A subgroup of celiac patients (≈ 5%) develops resistance to GFD
Association analysis strategy in RCDII in GWAS dataset

Discovery cohort
Dutch cohort
RCDII cases, n = 38
Healthy controls, n= 846

Hap550k SNPs
MAF > 0.01
HWE P > 0.00001
Fisher’s exact test

Top 15 loci with P < 10-5

Replication cohorts
1. France: RCDII cases, n = 33
   Healthy controls, n= 787
2. Italy: RCDII cases, n = 8+
   Healthy controls, n = 543
3. Dutch: RCDII cases, n = 7

Courtesy Vinod Kumar
Manhattan plot: Association between RCDII and healthy controls at 550K SNPs

Chromosomes

Manhattan plot: Association between RCDII and healthy controls at 550K SNPs

Chromosomes

Courtesy Vinod Kumar
<table>
<thead>
<tr>
<th>CHR</th>
<th>SNP</th>
<th>A1</th>
<th>F_A</th>
<th>F_U</th>
<th>A2</th>
<th>P</th>
<th>OR</th>
<th>A1</th>
<th>F_A</th>
<th>F_U</th>
<th>A2</th>
<th>P</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>SNP1</td>
<td>T</td>
<td>0,447</td>
<td>0,200</td>
<td>C</td>
<td>2,00E-006</td>
<td>3,24</td>
<td>T</td>
<td>0,447</td>
<td>0,185</td>
<td>C</td>
<td>3,58E-07</td>
<td>3,58</td>
</tr>
<tr>
<td>10</td>
<td>SNP2</td>
<td>A</td>
<td>0,013</td>
<td>0,179</td>
<td>G</td>
<td>1,42E-005</td>
<td>16,36</td>
<td>A</td>
<td>0,013</td>
<td>0,186</td>
<td>G</td>
<td>8,07E-06</td>
<td>17,14</td>
</tr>
<tr>
<td>17</td>
<td>SNP3</td>
<td>C</td>
<td>0,263</td>
<td>0,089</td>
<td>T</td>
<td>1,54E-005</td>
<td>3,65</td>
<td>C</td>
<td>0,263</td>
<td>0,093</td>
<td>T</td>
<td>3,07E-05</td>
<td>3,47</td>
</tr>
<tr>
<td>13</td>
<td>SNP4</td>
<td>A</td>
<td>0,447</td>
<td>0,218</td>
<td>G</td>
<td>1,60E-005</td>
<td>2,91</td>
<td>A</td>
<td>0,447</td>
<td>0,221</td>
<td>G</td>
<td>2,00E-05</td>
<td>2,86</td>
</tr>
<tr>
<td>14</td>
<td>SNP5</td>
<td>G</td>
<td>0,513</td>
<td>0,277</td>
<td>A</td>
<td>2,35E-005</td>
<td>2,76</td>
<td>G</td>
<td>0,513</td>
<td>0,241</td>
<td>A</td>
<td>9,97E-07</td>
<td>3,33</td>
</tr>
<tr>
<td>5</td>
<td>SNP6</td>
<td>A</td>
<td>0,408</td>
<td>0,191</td>
<td>G</td>
<td>2,38E-005</td>
<td>2,92</td>
<td>A</td>
<td>0,408</td>
<td>0,208</td>
<td>G</td>
<td>1,51E-04</td>
<td>2,63</td>
</tr>
<tr>
<td>16</td>
<td>SNP7</td>
<td>A</td>
<td>0,026</td>
<td>0,192</td>
<td>C</td>
<td>3,21E-005</td>
<td>8,80</td>
<td>A</td>
<td>0,026</td>
<td>0,193</td>
<td>C</td>
<td>4,92E-05</td>
<td>8,83</td>
</tr>
<tr>
<td>14</td>
<td>SNP8</td>
<td>T</td>
<td>0,197</td>
<td>0,428</td>
<td>C</td>
<td>4,25E-005</td>
<td>3,04</td>
<td>T</td>
<td>0,197</td>
<td>0,431</td>
<td>C</td>
<td>3,90E-05</td>
<td>3,08</td>
</tr>
<tr>
<td>10</td>
<td>SNP9</td>
<td>T</td>
<td>0,224</td>
<td>0,073</td>
<td>G</td>
<td>4,78E-005</td>
<td>3,68</td>
<td>T</td>
<td>0,224</td>
<td>0,079</td>
<td>G</td>
<td>1,35E-05</td>
<td>3,36</td>
</tr>
<tr>
<td>11</td>
<td>SNP10</td>
<td>A</td>
<td>0,711</td>
<td>0,470</td>
<td>G</td>
<td>5,07E-005</td>
<td>2,76</td>
<td>A</td>
<td>0,711</td>
<td>0,475</td>
<td>G</td>
<td>8,09E-05</td>
<td>2,72</td>
</tr>
<tr>
<td>17</td>
<td>SNP11</td>
<td>T</td>
<td>0,684</td>
<td>0,447</td>
<td>C</td>
<td>5,19E-005</td>
<td>2,68</td>
<td>T</td>
<td>0,684</td>
<td>0,452</td>
<td>C</td>
<td>8,35E-05</td>
<td>2,62</td>
</tr>
<tr>
<td>15</td>
<td>SNP12</td>
<td>C</td>
<td>0,000</td>
<td>0,131</td>
<td>T</td>
<td>5,26E-005</td>
<td>0,00</td>
<td>C</td>
<td>0,000</td>
<td>0,129</td>
<td>T</td>
<td>5,08E-05</td>
<td>0,00</td>
</tr>
<tr>
<td>18</td>
<td>SNP13</td>
<td>A</td>
<td>0,013</td>
<td>0,161</td>
<td>G</td>
<td>6,06E-005</td>
<td>14,43</td>
<td>A</td>
<td>0,013</td>
<td>0,164</td>
<td>G</td>
<td>5,94E-05</td>
<td>14,73</td>
</tr>
<tr>
<td>2</td>
<td>SNP14</td>
<td>T</td>
<td>0,211</td>
<td>0,067</td>
<td>C</td>
<td>6,41E-005</td>
<td>3,73</td>
<td>T</td>
<td>0,211</td>
<td>0,066</td>
<td>C</td>
<td>6,20E-05</td>
<td>3,77</td>
</tr>
<tr>
<td>16</td>
<td>SNP15</td>
<td>T</td>
<td>0,079</td>
<td>0,271</td>
<td>C</td>
<td>8,09E-005</td>
<td>4,34</td>
<td>T</td>
<td>0,079</td>
<td>0,296</td>
<td>C</td>
<td>1,10E-05</td>
<td>4,90</td>
</tr>
</tbody>
</table>

1. False positive findings is a problem due to the GWAS design; => replication step required.

2. False negatives of association will be a major problem in small discovery samples, even if the effects are large.

Courtesy Vinod Kumar
## Replication cohorts

1. France: RCDII cases, n = 33  
   Healthy controls, n= 787  
2. Italy: RCDII cases, n = 8+  
   Healthy controls, n = 543  
3. Dutch: RCDII cases, n = 7+

<table>
<thead>
<tr>
<th>CHR</th>
<th>SNP</th>
<th>A1</th>
<th>F_A</th>
<th>F_U</th>
<th>A2</th>
<th>P_Fisher's Exact</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>SNP1</td>
<td>T</td>
<td>0.197</td>
<td>0.194</td>
<td>C</td>
<td>1.00E+00</td>
<td>1.016</td>
</tr>
<tr>
<td>10</td>
<td>SNP2</td>
<td>T</td>
<td>0.121</td>
<td>0.151</td>
<td>C</td>
<td>6.00E-01</td>
<td>0.777</td>
</tr>
<tr>
<td>17</td>
<td>SNP3</td>
<td>C</td>
<td>0.106</td>
<td>0.088</td>
<td>T</td>
<td>6.55E-01</td>
<td>1.235</td>
</tr>
<tr>
<td>13</td>
<td>SNP4</td>
<td>A</td>
<td>0.364</td>
<td>0.251</td>
<td>G</td>
<td>4.44E-02</td>
<td>1.703</td>
</tr>
<tr>
<td>14</td>
<td>SNP5</td>
<td>G</td>
<td>0.242</td>
<td>0.259</td>
<td>A</td>
<td>8.86E-01</td>
<td>0.917</td>
</tr>
<tr>
<td>5</td>
<td>SNP6</td>
<td>T</td>
<td>0.227</td>
<td>0.233</td>
<td>C</td>
<td>1.00E+00</td>
<td>0.967</td>
</tr>
<tr>
<td>16</td>
<td>SNP7</td>
<td>T</td>
<td>0.359</td>
<td>0.208</td>
<td>G</td>
<td>7.35E-03</td>
<td>2.139</td>
</tr>
<tr>
<td>14</td>
<td>SNP8</td>
<td>T</td>
<td>0.424</td>
<td>0.461</td>
<td>C</td>
<td>6.15E-01</td>
<td>0.863</td>
</tr>
<tr>
<td>10</td>
<td>SNP9</td>
<td>T</td>
<td>0.076</td>
<td>0.062</td>
<td>G</td>
<td>6.04E-01</td>
<td>1.231</td>
</tr>
<tr>
<td>11</td>
<td>SNP10</td>
<td>A</td>
<td>0.422</td>
<td>0.482</td>
<td>G</td>
<td>3.74E-01</td>
<td>0.786</td>
</tr>
<tr>
<td>17</td>
<td>SNP11</td>
<td>A</td>
<td>0.394</td>
<td>0.453</td>
<td>G</td>
<td>3.78E-01</td>
<td>0.785</td>
</tr>
<tr>
<td>15</td>
<td>SNP12</td>
<td>C</td>
<td>0.076</td>
<td>0.119</td>
<td>T</td>
<td>4.31E-01</td>
<td>0.61</td>
</tr>
<tr>
<td>18</td>
<td>SNP13</td>
<td>T</td>
<td>0.182</td>
<td>0.14</td>
<td>C</td>
<td>3.66E-01</td>
<td>1.36</td>
</tr>
<tr>
<td>2</td>
<td>SNP14</td>
<td>T</td>
<td>0.091</td>
<td>0.082</td>
<td>C</td>
<td>8.18E-01</td>
<td>1.12</td>
</tr>
<tr>
<td>16</td>
<td>SNP15</td>
<td>A</td>
<td>0.182</td>
<td>0.249</td>
<td>G</td>
<td>2.45E-01</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Courtesy Vinod Kumar
Regional association plot at SNP4 on chr 13

SNP 4: Meta-P = 2.85 \times 10^{-6}

Courtesy Vinod Kumar
1. The protein encoded by this gene is a member of the Kruppel-like zinc finger protein family.

2. It can repress expression of the AP-2 alpha gene by binding to a specific site in the AP-2 alpha gene promoter. (Activator protein-2 alpha (AP-2 alpha) is a developmentally-regulated transcription factor and important regulator of gene expression during vertebrate development and carcinogenesis).

In plan - Exome-arrays (Illumina) to detect coding variants associated with RCDII
General discussion and conclusions

- GWAS studies have identified 39 non-HLA loci for celiac disease
  - Common variants with small effect sizes (OR <<2)
  - Majority seem to affect gene regulation (cis-eQTL, miRNAs, epigenetics?)

- Co-occurrence of immune-related diseases can be explained by genetics. Opportunities:
  - Cross-disease meta-analysis between clinically distinct diseases
  - Increases our understanding of what determines ‘disease specificity’
  - May help to define ‘networks’ of genes that are functionally related

- Challenges for the near future:
  - To understand how genetic variation impacts molecular and cellular phenotypes (‘systems’ approaches)
  - Identify the remainder of the genetic factors (both common and rare)
  - Unravel gene-gene and gene-environment interactions

- RCDII – increasing the sample size. Exome sequencing
Acknowledgements celiac disease research

Collaborators
David van Heel
Maria Teresa Bardella
Donatella Barosani
Sjoerd Bakker
Thelma BK
Chris Cotsapas
Bozena Cukrowska
Carlo Catassi
Mark Daly
Luigi Greco
Luisa Mearin
Vandana Midah
Chris Mulder
Mihai Netea
Robert Plenge
Anna Rybak
Paivi Saavalainen
Sabyasachi Senapati
Ajit Sood
Rinse Weersma
Victorien Wolters
PreventCD Study Group

Cisca Wijmenga, Nick Bockett, Rodrigo Coutinho de Almeida, Cleo van Diemen, Patrick Dubois, Jingyuan Fu, Javier Gutierrez-Achury, Graham Heap, Barabra Hrdličková, Karen Hunt, Juha Karjalainen, Vanisha Mistry, Isis Ponce, Vinod Kumar, Harm-Jan Westra, Sebo Withoff, Lude Franke, Jihane Romanos, Sasha Zhernakova, Agata Szperl, Gosia Trynka, Noortje Festen
Thank you