33 new Crohn’s disease susceptibility genes and loci identified by the International IBD Genetics Consortium

Miles Parkes on behalf of IIBDGC
Acknowledgements

Andre Franke
A. Hillary Steinhart
Albert Cohen
Amir Levine
Andre Van Gossum
Anne Griffiths
Anne Phillips
Carl Anderson
Carsten Buning
Cathryn Edwards
Cecile Libioulle
Charlie Lees
Christopher Mathew
Cisca Wijmenga
Craig Mowat
Daan Hommes
Debby Laukens
Deborah D Proctor
Denis Franchimont
Derek P Jewell
Dermot McGovern
Edouard Louis
Frank Seibold
G Radford Smith
Grant Montgomery
Hakon Hakonarson
Ian Lawrance
Jack Satsangi
James Lee
Jean-Fred Colombel
Jean-Pierre Hugot
Jeff Barrett
Jeremy Sanderson
Jerry Rotter
John Mansfield
John Rioux
Jonas Halfvarson
Judy Cho
Julian Panés
Jürgen Glas
Kai Wang
Kent Taylor
Leif Torkvist
Lisa Simms
Marc Lemann
Marie Cottone
Mark Daly
Mark Silverberg
Maria Dubinsky
Martine De Vos
Mauro D'Amato
Michel Georges
Miguel Regueiro
Miquel Sans
Murray Barclay
Paul Rutgeerts
Phil Schumm
Pieter Stokkers
Rebecca Roberts
Renata D'Inca
Richard Duerr
Richard Gearnay
Rinse Weersma
Robert Baldassano
Salvator Cucchiara
Severine Vermeire
Stefan Schreiber
Stephan Brand
Stephan Targan
Steve Guthery
Steven R Brant
Subra Kugathasan
Tarig Ahmad
Ted Denson
Theodore Bayless
Thomas Balschun
Tim Florin
Vito Annese
William Newman
Yashoda Sharma

Funding Support from NIH, the Wellcome Trust, Giuliani SpA and Abbott Laboratories
Crohn’s disease: Pathogenesis

CD results from a dysregulated immune response to gut bacteria in genetically susceptible individuals.

Characterising the susceptibility genes should

1. Identify primary pathogenic pathways
2. Identify new targets for drug therapy
3. Help understand the environmental drivers
Genome Wide Association Scanning

• Genotype ≥0.5 million SNPs in thousands of individuals

• hypothesis-free, unbiased survey of genome for susceptibility loci
GWAS studies in Crohn’s disease

ARTICLES

Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls

The Wellcome Trust Case Control Consortium*

There is increasing evidence that genome-wide association (GWA) studies represent a powerful approach to the identification of genes involved in common human diseases. We describe a joint GWA study (using the Affymetrix GeneChip 500K Mapping Array Set) undertaken in the British population, which has examined ~2,000 individuals for each of 7 major diseases and a shared set of ~3,000 controls. Case-control comparisons identified 24 independent association signals at $P < 5 \times 10^{-7}$: 1 in bipolar disorder, 1 in coronary artery disease, 9 in Crohn’s disease, 3 in rheumatoid arthritis, 7 in type 1
GWAS studies in Crohn’s disease

- Identified importance of autophagy and Th17 pathways etc.
1st International Crohn’s disease GWAS Meta-analysis: 32 confirmed loci
Barrett et al. *Nature Genetics* 2008

Genome-wide association defines more than 30 distinct susceptibility loci for Crohn’s disease
But....

• Only 20% of genetic variance explained...
Aim

• Identify additional Crohn’s disease susceptibility genes and loci which surpass a stringent genome-wide significance threshold of $P < 5 \times 10^{-8}$
Methods - overview

• meta-analysis of all available Crohn’s disease GWAS’d case-controls (‘discovery panel’)

• Follow-up new signals in independent replication panel

NB: expanded ++ consortium has much greater power to detect ‘typical’ loci of modest effect
Subjects: GWAS discovery panel for meta-analysis

<table>
<thead>
<tr>
<th>Index GWAS</th>
<th>Crohn’s disease cases</th>
<th>Healthy controls</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adolescent</td>
<td>1689</td>
<td>6197</td>
<td>Nat Genet 2009</td>
</tr>
<tr>
<td>German</td>
<td>479</td>
<td>1145</td>
<td>PLoS Genet 2009</td>
</tr>
<tr>
<td>USA (Cedars Sinai)</td>
<td>925</td>
<td>2882</td>
<td></td>
</tr>
<tr>
<td>Belgium</td>
<td>537</td>
<td>913</td>
<td>PLoS Genet 2007</td>
</tr>
<tr>
<td>USA (NIDDK)</td>
<td>956</td>
<td>982</td>
<td>Nature Genet 2007</td>
</tr>
<tr>
<td>UK (WTCCC)</td>
<td>1747</td>
<td>2937</td>
<td>Nature 2007</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>6,333</strong></td>
<td><strong>15,056</strong></td>
<td></td>
</tr>
</tbody>
</table>
Methods: meta-analysis

- Used GWAS data +‘Beagle’ to impute genotypes for 953,000 HapMap3 SNPs / sample
- Standard 1 df allele-based tests of association – summarized as Z scores & p values for each SNP
- Each new locus meeting $P<1 \times 10^{-5}$ => ‘focal’ SNP +/- proxy genotyped in large replication panel
Subjects: replication panel

<table>
<thead>
<tr>
<th>REPLICATION</th>
<th>cases</th>
<th>controls</th>
<th>platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>1357</td>
<td>1923</td>
<td>Sequenom</td>
</tr>
<tr>
<td>Belgium</td>
<td>1282</td>
<td>1682</td>
<td>SNPlex/Taqman</td>
</tr>
<tr>
<td>France</td>
<td>414 trios</td>
<td></td>
<td>SNPlex/Taqman</td>
</tr>
<tr>
<td>Germany</td>
<td>3808</td>
<td>2747</td>
<td>SNPlex/Taqman</td>
</tr>
<tr>
<td>Israel</td>
<td>444</td>
<td>376</td>
<td>SNPlex/Taqman</td>
</tr>
<tr>
<td>Italy</td>
<td>921</td>
<td>899</td>
<td>SNPlex/Taqman</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1101</td>
<td>269</td>
<td>SNPlex/Taqman</td>
</tr>
<tr>
<td>New Zealand</td>
<td>514</td>
<td>457</td>
<td>Sequenom</td>
</tr>
<tr>
<td>Spain</td>
<td>325</td>
<td>987</td>
<td>SNPlex/Taqman</td>
</tr>
<tr>
<td>Sweden</td>
<td>724</td>
<td>992</td>
<td>SNPlex/Taqman</td>
</tr>
<tr>
<td>UK</td>
<td>3243</td>
<td>2431</td>
<td>Sequenom</td>
</tr>
<tr>
<td>USA (cedars)</td>
<td>1172</td>
<td>501</td>
<td>Sequenom</td>
</tr>
<tr>
<td>USA (NIDDK)</td>
<td>803</td>
<td>762</td>
<td>Illumina</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>14,934</strong></td>
<td><strong>13,647</strong></td>
<td>****</td>
</tr>
</tbody>
</table>
Results

- 52 new loci met $P<1 \times 10^{-5}$ in discovery panel $\Rightarrow$ focal SNP genotyped in replication panel
- SNPs from 33 distinct new loci showed evidence of association at $P<5 \times 10^{-8}$ in combined ‘discovery + replication’ panel (with $P<0.05$ in replication)
Representative association plot

‘one gene’ locus
Representative association plot

‘Gene desert’
Representative association plot

‘Multi-gene’ locus
### Loci with a single gene

<table>
<thead>
<tr>
<th>SNP</th>
<th>CHR</th>
<th>meta-GWAS P</th>
<th>Rep P</th>
<th>Combined P</th>
<th>OR</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs17293632</td>
<td>15</td>
<td>1.41E-13</td>
<td>2.00E-08</td>
<td><strong>2.70E-19</strong></td>
<td>1.12</td>
<td>SMAD3</td>
</tr>
<tr>
<td>rs212388</td>
<td>6</td>
<td>1.41E-07</td>
<td>2.40E-05</td>
<td><strong>2.30E-11</strong></td>
<td>1.1</td>
<td>TAGAP</td>
</tr>
<tr>
<td>rs13428812</td>
<td>2</td>
<td>1.41E-08</td>
<td>0.00059</td>
<td><strong>8.50E-10</strong></td>
<td>1.06</td>
<td>DNMT3A</td>
</tr>
<tr>
<td>rs11167764</td>
<td>5</td>
<td>1.1E-09</td>
<td>0.0042</td>
<td><strong>2.00E-09</strong></td>
<td>1.06</td>
<td>NDFIP1</td>
</tr>
<tr>
<td>rs12722489</td>
<td>10</td>
<td>8.51E-06</td>
<td>5.20E-05</td>
<td><strong>2.90E-09</strong></td>
<td>1.11</td>
<td>IL2RA</td>
</tr>
<tr>
<td>rs1847472</td>
<td>6</td>
<td>3.63E-06</td>
<td>1.40E-04</td>
<td><strong>5.10E-09</strong></td>
<td>1.07</td>
<td>BACH2</td>
</tr>
<tr>
<td>rs1998598</td>
<td>1</td>
<td>4.9E-09</td>
<td>0.016</td>
<td><strong>8.70E-09</strong></td>
<td>1.04</td>
<td>DENND1B</td>
</tr>
</tbody>
</table>
SMAD3

- Transcriptional regulator downstream of TGF-β
- key role in TGF-β-mediated induction of Foxp3⁺ Tregs
- SMAD3 deficiency => incr Th17 differentiation
- Reduced SMAD3 phosphorylation seen in IBD
  - Tone et al. *Nat Immunol* 2008
  - Lu et al. *J Immunol* 2010
TAGAP

- T-cell Activation GTPase-Activating Protein
- Regulator of T cell activation / co-regulated with IL2
- Also associated with
  - celiac disease
  - rheumatoid arthritis
  - type 1 diabetes
Loci with a single gene

**DNMT3A**

- encodes DNA methyltransferase 3a - one of three key methyltransferases
- epigenetic regulator of gene transcription by methylating cytosine in CpG islands.
- many roles incl dynamic regulation of innate + adaptive immunity

Loci with a single gene

**IL2RA**
- encodes part of IL2 receptor
- regulates response to autoantigens by modulating Foxp3$^+$ Tregs
- associated SNP correlates with IL2RA (CD25) expression on CD4+ naïve + memory T cells
- also associated with T1D + MS
  - Dendrou et al *Nature Genetics* 2009
Loci with a single gene

DENND1B

– expressed in dendritic and effector memory T cells
– represses TNFR1 signaling => modulates Th1-Th2 cytokines
– Recent assoc with asthma

– Sleiman et al *NEJM* 2010
Methods to identify key gene within multi-gene loci

- eQTL analysis
  - to correlate focal SNP variation with gene expression
- Coding SNP (cSNP) analysis
  - using 1000G and Hapmap3 to identify genes containing cSNP’s in LD with focal SNP
- GRAIL (pathway) analysis
  - Gene Relationships Among Implicated Loci
  - identify functional connectivity between genes in CD-associated loci
GRAIL connectivity for 69 CD loci
### ‘One from Many’ Analysis

<table>
<thead>
<tr>
<th>SNP</th>
<th>CHR</th>
<th>meta-GWAS P</th>
<th>Rep P</th>
<th>Combined P</th>
<th>OR</th>
<th>Gene Highlighted</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs281379</td>
<td>19</td>
<td>8.60E-10</td>
<td>5.20E-05</td>
<td>7.40E-12</td>
<td>1.07</td>
<td>FUT2</td>
</tr>
<tr>
<td>rs2549794</td>
<td>5</td>
<td>4.47E-11</td>
<td>0.002</td>
<td>1.10E-10</td>
<td>1.05</td>
<td>ERAP2</td>
</tr>
<tr>
<td>rs12720356</td>
<td>19</td>
<td>9.20E-10</td>
<td>1.90E-05</td>
<td>1.40E-12</td>
<td>1.12</td>
<td>TYK2</td>
</tr>
<tr>
<td>rs3024505</td>
<td>1</td>
<td>8.32E-09</td>
<td>1.50E-07</td>
<td>1.60E-14</td>
<td>1.12</td>
<td>IL10</td>
</tr>
</tbody>
</table>
One from Many

**FUT2** (cSNP)

- encodes alpha-(1,2)fucosyltransferase
- regulates expression of Lewis ABO(H) blood group antigens on epithelial cell surface
- 20% population *FUT2* non-secretor genotype => marked resistance to norovirus
- also partial resistance to H pylori and HIV infection
ERAP2 (strong eQTL)

- endoplasmic reticulum aminopeptidase 2
- regulated by NFκB
- trims peptides for presentation on MHC class I => critically affects antigen presentation to T cells
- NB: ERAP1 associated with ank spond
  - Saveanu et al *Nature Immunology* 2005
TYK2 (cSNP + GRAIL)

- encodes tyrosine kinase 2 (JAK family)
- transduces $\gamma$-IFN, IL12 and IL23 signaling
  => regulates Th1 / Th17 development
- also role in TLR-mediated responses in dendritic cells, incl IL12 + IL23 production
  => Ghoreschi et al Immunol Rev. 2009
Gene desert

**TNFSF11 = RANKL (eQTL)**

- receptor activator of NFκB
- encodes a member of TNF family
- stimulates dendritic cells => proliferation of naive + systemic Treg populations
- also regulates osteoclast activity / bone loss
- previous studies => increased plasma levels in CD
  - Moschen et al. *Gut* 2005
What proportion of genetic variance now explained?

~25%
Conclusions

• 33 new Crohn’s disease susceptibility loci identified
  • Total tally confirmed CD loci now = 69
• Specific innate and adaptive immune pathways highlighted
• Also role of
  – epigenetic regulation (DNMT3a)
  – pathogen resistance (FUT2)
• Fine mapping in progress to identify causal variants