Inflammatory bowel disease: ulcerative colitis and Crohn’s disease

- Symptoms: diarrhea, abdominal pain, intestinal bleeding, growth retardation
- Pathophysiology: ubiquitous, commensal intestinal bacteria trigger an overactive and ongoing mucosal immune response in genetically predisposed hosts
- Prevalence: 100-200/100,000 for each disorder
- Peak age of onset: 2nd to 4th decades of life
Lines of evidence for heritability of IBD

- Racial and ethnic differences in rates of IBD
  - Whites > Blacks > Hispanics and Asians in U.S.
  - Jews > non-Jews

- Familial aggregation of IBD
  - 5-20% of patients have a family hx of IBD
  - $\lambda_s$ as high as 30-40 for CD and 10-20 for UC
  - 75% of multiplex families have only CD- or only UC-affected members; 25% of families “mixed” with CD and UC within the same family

- MZ: DZ twin concordance
  - CD: MZ=20-50% : DZ=0-7%
  - UC: MZ=14-19% : DZ=0-5%

- Associations between IBD and specific genetic variants
Confirmed Crohn’s disease loci

- **NOD2**
- **IBD5**

Prior to GWAS
Confirmed Crohn’s disease loci

- **NOD2**
- **IBD5**
- **TNFSF15**

- Prior to GWAS
- Early GWAS
Genetic Research Centers

Cedars – Sinai Medical Center
- Kent Taylor (PI)

Johns Hopkins University
- Steven Brant (PI)

Universite de Montreal
- John Rioux (PI)

University of Pittsburgh
- Richard Duerr (PI)

University of Toronto
- Mark Silverberg (PI)

Yale University
- Judy Cho (PI)

Data Coordinating Center

Yale University
- Judy Cho (PI), Yashoda Sharma

University of Chicago
- Dan Nicolae, Philip Schumm, Emily Kistner

Committees

Recruitment - Mark Silverberg
Phenotyping - Hillary Steinhart
Genotyping - Richard Duerr
Analytic - Mark Daly
Publications/IP - Steven Brant, John Rioux

A Genome-Wide Association Study Identifies IL23R as an Inflammatory Bowel Disease Gene

Genome-wide association study identifies new susceptibility loci for Crohn’s disease and implicates autophagy in disease pathogenesis

GWAS Crohn’s cases: N=946
GWAS controls: N=977
Platform: Illumina HumanHap300
NOD2
IL23R
ATG16L1 10q21

genome-wide significance (Bonferroni corrected \( p < 0.05 \))

significance threshold chosen for further follow-up studies
Belgian-French IBD consortium

- Teams:
  - University of Liège: C. Libioulle, S. Hansoul, C. Sandor, M. Mni, J. Belaiche, E. Louis & M. Georges
  - Catholic University of Leuven: S. Vermeire, P. Rutgeerts
  - Free University of Brussels: D. Franchimont, A. Van Gossum
  - Ghent University: M. De Vos, D. Laukens
  - University Paris Diderot: J.P. Hugot

- Samples:
  - GWAS Crohn's cases: N=547
  - GWAS controls: N=928
  - Platform: Illumina HumanHap300

Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4
CD-associated chromosome 5p13 variants and *PTGER4* eQTL analysis in EBV-transformed lymphoblastoid cell lines

Confirmed Crohn’s disease loci

- **NOD2**
- **IBD5**
- **TNFSF15**
- **IL23R**
- **ATG16L1**
- **10q21**
- **5p13**

- Prior to GWAS
- Early GWAS
NIDDK IBDGC Crohn's disease GWA study

- log10(P)

NOD2
IL23R
ATG16L1
5p13
10q21
IBD5
Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn’s disease susceptibility.

Miles Parkes1,3, Jeffrey C Barrett2,13, Natalie J Prescott3,13, Mark Tremelling1, Carl A Anderson1, Sheila A Fisher1, Roland G Roberts3, Elaine R Nimmo4, Fraser R Cummings3, Dianne Soars3, Hazel Drummond4, Charlie W Lees4, Saud A Khawaja3, Richard Bagnall3, Denis A Burke6, Catherine E Todhunter6, Tariq Ahmad5, Clive M Omnie3, Wendy McArdle7, David Strachan8, Graeme Bethel9, Claire Bryan9, Cathryn M Lewis8, Panos Deloukas8, Alastair Forbes10, Jeremy Sanderson11, Derek P Jewell5, Jack Satsangi8, John C Mansfield6, the Wellcome Trust Case Control Consortium13, Lon Cardon7 & Christopher G Mathew3

GWAS Crohn’s cases: N=1,748
GWAS controls: N=2,938 (’58 BBC/NBS donors)
Platform: Affymetrix 500K
WTCCC Crohn's disease GWA study

- NOD2
- ATG16L1
- IL23R
- 5p13
- IBD5
- IRGM
- 10q21
- NKX2-3
- PTEN2
- 3p21 (many genes)

Many genes are associated with Crohn's disease in this study.
Confirmed Crohn’s disease loci

- NOD2
- IBD5
- TNFSF15
- IL23R
- 3p21
- IRGM
- NKX2-3
- PTPN2
- ATG16L1
- 10q21
- 5p13

- Prior to GWAS
- WTCCC GWAS
- Early GWAS
Opportunity for additional disease locus discovery: Crohn’s disease GWA meta-analysis

- meta-analysis of NI DDK + Belgian/French + WTCCC Crohn’s disease GWA studies
- combined total of 3,230 cases and 4,829 controls with genome-wide SNP data post quality control
- 633,548 SNPs (directly genotyped SNPs or SNPs with genotypes imputed using genotyped SNPs and HapMap data)
Crohn’s disease GWA meta-analysis provides substantial increase in discovery power

(Example – disease-associated SNP with MAF = 0.20)
74 independent regions had at least one SNP with $p < 5 \times 10^{-5}$ (blue line). These regions were targeted in a further replication study.
Crohn’s disease GWA meta-analysis replication

- Replication study subjects, post quality control
  - 2,325 cases
  - 1,809 controls
  - 1,339 parent-parent-affected offspring trios
- 30 regions had significant replication evidence and had combined (GWA meta-analysis and replication) genome-wide significant $p < 5 \times 10^{-8}$
- Another 10 regions were nominally replicated with $p < 0.05$ (would expect only 2-3 by chance); the MHC and chr 19p13 had combined (GWA meta-analysis and replication) genome-wide significant $p < 5 \times 10^{-8}$
Confirmed Crohn’s disease loci

- ~ 10% overall risk
- ~ 20% genetic risk

~ 10% overall risk
~ 20% genetic risk

prior to GWAS
WTCCC GWAS
early GWAS
GWAS meta-analysis
IL23/TH17 pathway variants associated with risk for CD

**IL23R Arg381Gln**

Gln allele is protective

*Frequency: ~2% in cases ~7% in controls*

***Also associated with UC and psoriasis***

**IL23R risk haplotype**

Causal variant not known

*Frequency: ~40% in cases ~30% in controls*

---

Duerr et al. Science 2006;314:1461-3

Autophagy in pathogen degradation

- Pathogen in intracellular compartment
- Pathogen degradation in lysosome
- Pathogen in autophagosome

Autophagy in antigen presentation

- Chaperone mediated autophagy
- Loading of MHC class II with autophagosome derived peptide
- CD4 - MHC class II
- CLIP
Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease


- scan of 10,886 nsSNPs in 905 UC cases and 1,465 controls (panel 1)
- 33 SNPs from 21 distinct loci with P < 0.001 were followed up in 936 UC cases and 1,470 controls (panel 2)
- 2 SNPs with promising association evidence in the ECM1 (extracellular matrix protein 1) locus, which had not previously been associated with IBD, were followed up in an additional 1,146 cases and 1,559 controls (panel 3)
- 16 SNPs tagging 13 Crohn’s disease–associated loci identified by GWAS were tested in 1,841 ulcerative colitis cases (panels 1 and 2) and 1,470 controls (panel 2)
- association analyses using Cochran-Armitage trend tests
Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease


<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Pos</th>
<th>Gene</th>
<th>PGC panel 1</th>
<th>P panel 2</th>
<th>P panel 3</th>
<th>P combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3737240</td>
<td>1</td>
<td>148749979</td>
<td>ECM1</td>
<td>4.5 x 10^-4</td>
<td>0.013</td>
<td>0.038</td>
<td>2.3 x 10^-6</td>
</tr>
<tr>
<td>rs13294</td>
<td>1</td>
<td>148751611</td>
<td>ECM1</td>
<td>7.1 x 10^-4</td>
<td>0.020</td>
<td>0.060</td>
<td>7.9 x 10^-6</td>
</tr>
<tr>
<td>rs3197999</td>
<td>3</td>
<td>49696536</td>
<td>MST1</td>
<td>0.0010</td>
<td>0.0061</td>
<td></td>
<td>8.0 x 10^-6</td>
</tr>
<tr>
<td>rs9268480</td>
<td>6</td>
<td>32471822</td>
<td>BTNL2</td>
<td>0.0015</td>
<td>0.0064</td>
<td></td>
<td>7.2 x 10^-6</td>
</tr>
<tr>
<td>rs660895</td>
<td>6</td>
<td>32685358</td>
<td>HLA-DRB1</td>
<td>1.5 x 10^-5</td>
<td>0.0035</td>
<td></td>
<td>2.8 x 10^-8</td>
</tr>
</tbody>
</table>
Genetic determinants of ulcerative colitis include the *ECM1* locus and five loci implicated in Crohn's disease.


<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Pos</th>
<th>Gene</th>
<th>P combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11805303</td>
<td>1</td>
<td>67448104</td>
<td><em>IL23R</em></td>
<td>2.2 x 10^{-4}</td>
</tr>
<tr>
<td>rs9858542</td>
<td>3</td>
<td>49676987</td>
<td><em>MST1</em></td>
<td>1.3 x 10^{-6}</td>
</tr>
<tr>
<td>rs9292777</td>
<td>5</td>
<td>40473705</td>
<td>gene desert</td>
<td>0.014</td>
</tr>
<tr>
<td>rs6556416</td>
<td>5</td>
<td>158751323</td>
<td><em>IL12B</em></td>
<td>6.8 x 10^{-4}</td>
</tr>
<tr>
<td>rs6887695</td>
<td>5</td>
<td>158755223</td>
<td><em>IL12B</em></td>
<td>0.0016</td>
</tr>
<tr>
<td>rs10761659</td>
<td>10</td>
<td>64115570</td>
<td>gene desert</td>
<td>0.0012</td>
</tr>
<tr>
<td>rs10883365</td>
<td>10</td>
<td></td>
<td><em>NKX2-3</em></td>
<td>2.4 x 10^{-6}</td>
</tr>
</tbody>
</table>
Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility


- scan of 440,794 SNPs in 1,167 UC cases and 777 controls (panel A)
- top 20 SNPs were followed up in three case-control samples totalling 1,855 UC cases and 3,091 controls in stage 2 (panels B-D)
Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility


<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Pos</th>
<th>Gene</th>
<th>Panel A</th>
<th>Panels B-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11805303</td>
<td>1</td>
<td>67387537</td>
<td>IL23R</td>
<td>5.39 x 10^{-6}</td>
<td>1.09 x 10^{-5}</td>
</tr>
<tr>
<td>rs3024505</td>
<td>1</td>
<td>203328299</td>
<td>IL10</td>
<td>1.43 x 10^{-5}</td>
<td>1.35 x 10^{-12}</td>
</tr>
<tr>
<td>rs12612347</td>
<td>2</td>
<td>158755223</td>
<td>ARPC2</td>
<td>8.42 x 10^{-6}</td>
<td>2.00 x 10^{-4}</td>
</tr>
<tr>
<td>rs9268480</td>
<td>6</td>
<td>32471822</td>
<td>BTNL2</td>
<td>2.21 x 10^{-6}</td>
<td>3.15 x 10^{-9}</td>
</tr>
<tr>
<td>rs9268858</td>
<td>6</td>
<td>32537736</td>
<td>HLA-DRA</td>
<td>5.41 x 10^{-7}</td>
<td>2.58 x 10^{-12}</td>
</tr>
<tr>
<td>rs9268877</td>
<td>6</td>
<td>32539125</td>
<td>HLA-DRA</td>
<td>5.23 x 10^{-7}</td>
<td>6.48 x 10^{-18}</td>
</tr>
</tbody>
</table>
Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study

**NI DDK I BDGC UC GWAS stage I**

<table>
<thead>
<tr>
<th>Study subjects</th>
<th>Genotyping array</th>
<th>Raw data (number of samples)</th>
<th>Post quality control data (number of samples)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ulcerative colitis cases</td>
<td>Illumina HumanHap300v2</td>
<td>552</td>
<td>537</td>
</tr>
<tr>
<td>Ulcerative colitis cases</td>
<td>Illumina HumanHap550v3</td>
<td>500</td>
<td>485</td>
</tr>
<tr>
<td>Total ulcerative colitis cases</td>
<td>1,052</td>
<td>1,022</td>
<td></td>
</tr>
<tr>
<td>NIDDK I BDGC Crohn's disease GWAS controls</td>
<td>Illumina HumanHap300v1</td>
<td>994</td>
<td>951</td>
</tr>
<tr>
<td>Illumina iControlDB study 65 controls</td>
<td>Illumina HumanHap300v1</td>
<td>125</td>
<td>123</td>
</tr>
<tr>
<td>Illumina iControlDB study 65 controls</td>
<td>Illumina HumanHap550v1</td>
<td>619</td>
<td>609</td>
</tr>
<tr>
<td>Illumina iControlDB studies 64 and 65 controls</td>
<td>Illumina HumanHap550v3</td>
<td>833</td>
<td>820</td>
</tr>
<tr>
<td>Total controls</td>
<td>2,571</td>
<td>2,503</td>
<td></td>
</tr>
</tbody>
</table>
Genetic matching (GEM) to control for population structure in case-control association analysis

- create an $L$ SNPs and $N$ individuals matrix of allele counts from genome-wide tag SNP data
- compute eigenvector decomposition to describe the ancestry of individuals on a multidimensional “principal components of ancestry” map
- remove outliers for any significant principal component of ancestry and recompute eigenvector decomposition
- cluster individuals who appear to have common ancestry; remove individuals who are too distant from their counterparts (controls for cases, and vice versa)
- perform association analysis using cluster number as the conditional variable in conditional logistic regression analysis
Principal components of ancestry in NIDDK IBDGC UC GWAS stage I female study subject tag SNP data.
### NI DDK IBDGC UC GWAS stage I

<table>
<thead>
<tr>
<th>Study subjects</th>
<th>Genotyping method</th>
<th>All arrays shared SNPs</th>
<th>HumanHap550 only SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>gender- and ancestry-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>matched case-control</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>strata (number of</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>samples, strata or</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SNPs)</td>
<td></td>
</tr>
<tr>
<td>Ulcerative colitis cases</td>
<td>Illumina HumanHap300v2</td>
<td>512</td>
<td></td>
</tr>
<tr>
<td>Ulcerative colitis cases</td>
<td>Illumina HumanHap550v3</td>
<td>465</td>
<td>476</td>
</tr>
<tr>
<td>Total ulcerative colitis cases</td>
<td></td>
<td>977</td>
<td>476</td>
</tr>
<tr>
<td>NI DDK IBDGC Crohn's disease GWAS controls</td>
<td>Illumina HumanHap300v1</td>
<td>837</td>
<td></td>
</tr>
<tr>
<td>Illumina iControlDB study 65 controls</td>
<td>Illumina HumanHap300v1</td>
<td>102</td>
<td></td>
</tr>
<tr>
<td>Illumina iControlDB study 65 controls</td>
<td>Illumina HumanHap550v1</td>
<td>523</td>
<td>565</td>
</tr>
<tr>
<td>Illumina iControlDB studies 64 and 65 controls</td>
<td>Illumina HumanHap550v3</td>
<td>660</td>
<td>729</td>
</tr>
<tr>
<td>Total controls</td>
<td></td>
<td>2,122</td>
<td>1,294</td>
</tr>
<tr>
<td>Gender- and ancestry-matched case-control strata</td>
<td></td>
<td>909</td>
<td>447</td>
</tr>
<tr>
<td>SNPs</td>
<td></td>
<td>280,748</td>
<td>207,332</td>
</tr>
</tbody>
</table>
NI DDK I BDGC UC GWAS stage I
quantile-quantile plots pre- and post-GEM

(a)

(b)

pre-GEM genomic control inflation factor = 1.17
post-GEM genomic control inflation factor = 1.04
**NI DDK I BDGC UC GWAS stage II**

<table>
<thead>
<tr>
<th>Study subjects</th>
<th>Raw data (number of samples)</th>
<th>Post quality control data (number of samples)</th>
</tr>
</thead>
<tbody>
<tr>
<td>North American cases</td>
<td>769</td>
<td>768</td>
</tr>
<tr>
<td>North American controls</td>
<td>727</td>
<td>721</td>
</tr>
<tr>
<td>Italian cases</td>
<td>633</td>
<td>619</td>
</tr>
<tr>
<td>Italian controls</td>
<td>415</td>
<td>394</td>
</tr>
</tbody>
</table>

- Successfully genotyped 54 of 65 independent ($r^2 < 0.5$) SNPs with $P < 1 \times 10^{-4}$ in stage I
- Cochran-Mantel-Haenszel analysis of North American and Italian stage II case-control data
- Stage I and stage II $P$ values combined using Fisher method
rs10889677
P = 1.3 x 10^{-8}

rs11209026
P = 1.3 x 10^{-8}

rs1004819
P = 1.5 x 10^{-6}

rs11465804
P = 9.4 x 10^{-7}

rs1004819
P = 1.5 x 10^{-6}

rs11465804
P = 9.4 x 10^{-7}

rs2395185
P = 1.0 x 10^{-16}
UC GWAS Conclusions

- Loci identified in UC GWAS and confirmed in independent samples
  - chromosome 1p36 near **OTUD3** (OTU domain containing 3)  
    **PLA2G2E** (phospholipase A2, group IIE), and **RNF186** (ring finger protein 186)
  - **IL23R** on chromosome 1p31
  - chromosome 1q32 near **IL10**
  - MHC on chromosome 6p21
  - chromosome 12q15 near **IFNG, IL26, IL22**
- Similar to the experience in CD, more UC loci will likely be found as additional UC GWAS and a UC GWAS meta-analysis are done
What next?

- Resequencing to identify low frequency variants
- “Fine mapping” including trans-racial/trans-ethnic mapping
- Copy number variants
- Multimarker model correlation with subphenotypes/disease outcomes
- Functional studies—GWAS findings have stimulated a resurgence of biological investigation!