Epstein-Barr Virus in Autoimmune & Inflammatory Diseases: a new genomic perspective

John B. Harley, MD, PhD
Professor of Pediatrics & Medicine
The Center for Autoimmune Genomics and Etiology
Cincinnati Children’s Hospital Medical Center
University of Cincinnati College of Medicine

Microbiol Threats Workshop
National Academy of Sciences
Rockefeller Foundation, 420 Fifth Ave, New York city
June 11 & 12, 2019
First image of a lupus malar rash from 1856.

von Hebra
(Chair of Dermatology, Vienna)

(von Hebra. Atlas of Skin Diseases, 1856-1876)
The critical moment

EBNA-1 = Epstein–Barr nuclear antigen 1.

Heteroimmune antibodies

Ig maturation

EBNA-1

Self

The first autoantibody

Ig maturation

Self

Autoantibodies

Harley & James Bull NYU Hosp Joint Dis 64:45, 2006
Epstein-Barr virus (EBV)

- A γ herpes virus
- Transmission by saliva and nursing
- Infects epithelial & B cells. Some NK, T, macrophages
- Latent phase avoids and suppresses immune responses.
Diseases attributed to Epstein-Barr Virus (EBV)

- **Asymptomatic** or **Upper respiratory illness**. >90% adults are infected.
- **Mononucleosis** – common in teenagers (~3%), ♀>♂
- **Multiple sclerosis** – virtually all cases, epidemiologic evidence
- **Nasopharyngeal cancer** – nearly all cases. Southeast Asia. Latency II
- **Hodgkin’s lymphoma** - ~1/3 of cases. Latency II
- **Carcinoma of the stomach** – 10% of cases. Latency II
- **Diffuse large B cell lymphoma** activated cell type – immune compromise Latency III>II
- **Immunoblastoid B cell lymphoma** – immune compromise. Latency III
- **Chronic active EBV** – rare. Often with genetic defects.
- **Leiomyosarcomas** – rare - immune compromise.
- **T & NK lymphomas** – rare. Latency II
- **Burkitt’s lymphoma** – rare outside of Africa. Latency I
**Strategy for a Genome-Wide Association Study (GWAS).**

\[(\text{red}=T \text{ and black}=C)\]

Metanalysis of lupus genetic associations

Why do we find these associations? What do they mean?

Strategic goal

Genetics
Sex; Ancestry
> 100 genes

Environment
Epstein–Barr virus and immune history

Unified understanding of disease etiology

...better Diagnosis, Therapeutics, Prevention & Prognosis!
Questions:

• If EBV causes lupus then how do we find the mechanism?

Assume:
Both the genes and the environment contribute to the origin(s) of lupus.

Observe:
- The genetic identify changes in allelic frequency with no insight into mechanism.
- 90% of genetic loci are predicted to be regulatory
- Transcription factors are regulators of gene expression
Assume:
Both the genes and the environment contribute to the origin(s) of lupus.

Observe:
- The genetics identify changes in allelic frequency with no insight into mechanism.
- 90% of genetic loci are predicted to be regulatory
- Transcription factors are regulators of gene expression

Prediction:
- If the assumption and observation are true, then we predict that the transcription factors made by EBV are concentrated in the lupus genetic loci as a test of the Hypothesis that EBV causes SLE.
Chromatin Immunoprecipitation followed by next generation DNA Sequencing.

The antibody ("★") binding the transcription factor is the most important reagent.

Szalkowski AM, Briefings in Bioinform, 2010.
Systemic Lupus Erythematosus

Systemic lupus erythematosus in European Ancestry

50 lupus-associated genomic loci for SLE

TF, transcription factor.
EBNA-2 from EBV immunoprecipitates variants in 26 of 53 lupus risk loci (European ancestry) in EBV-transformed B cells, while EBNA-1 does not interact with any

EBNA-2, found at 26 of 53 lupus loci in Mutu cells, by simulation (RELI) is enriched with RR = 6.0 and a corrected probability of $1.1 \times 10^{-25}$

RR, relative risk.

Harley, et al
Nat Genet, 2018

**SLE:** 66 TFs $P_c < 10^{-6}$

1544 TF datasets tested

47/66 (>70%) are EBV tnsf B cells

50/66 (>75%) B cell origin
SLE loci more strongly intersect TF ChIP-seq peaks in EBV-infected B-cell lines compared with uninfected B-cell lines

Comparison of EBV-infected B cell lines (blue bars) with EBV-negative B cells (green bars). The y-axis shows the distribution of the RELI $-\log(p_{c}s)$ for each of the eight TFs with available data.

Nearly all of the TF ChIP-seq datasets associations with lupus risk loci are found in EBV-transformed B-cell lines (blue bars), when compared with datasets with that TF in any other cell type (green bars).

SLE: $P_c < 10^{-6}$

Red TFs from EBV super-enhancers

White matter hyperintensities on magnetic resonance imaging (axial fluid attenuated inversion recovery sequence) in two 80 year old patients: (left) minor white matter hyperintensities; (right) extensive white matter hyperintensities predominating in periventricular region.

Debette S & Markus HS. BMJ 2010;341:bmj.c3666
Rheumatoid Arthritis

From: http://gederodi.blogspot.com/2015/09/artritis-reumatoide.html

From: https://kidsgetarthritis.wordpress.com/2010/07/06
**Type 1 Diabetes**

**Healthy**
- Insulin moves glucose to cells
- Glucose enters cells
- Pancreas produces insulin

**Diabetic**
- More glucose in the blood
- Pancreas cannot produce insulin
- Immune cells destroy beta cells in the pancreas
### Viral serologies in new onset T1D versus controls in 23 viruses.

42 new onset T1D and 42 matched controls tested against 646 viral open reading frames from 23 viruses generated *ex vivo*, spotted onto glass and tested for antibody binding.

<table>
<thead>
<tr>
<th>Virus</th>
<th>T1D</th>
<th>Controls</th>
<th>Odds</th>
<th>95% CI</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>(control)</td>
<td># (%)</td>
<td># (%)</td>
<td>Ratio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMV cytomegalovirus</td>
<td>24 (57)</td>
<td>23 (55)</td>
<td>1.1</td>
<td>0.4-2.8</td>
<td>1</td>
</tr>
<tr>
<td>EBV Epstein-Barr virus</td>
<td>37 (88)</td>
<td>22 (52)</td>
<td>6.6</td>
<td>2.0-26</td>
<td>0.018</td>
</tr>
<tr>
<td>CVB coxsackievirus</td>
<td>42 (100)</td>
<td>42 (100)</td>
<td>0</td>
<td>0-∞</td>
<td>1</td>
</tr>
<tr>
<td>RUBA rubella virus</td>
<td>22 (52)</td>
<td>25 (60)</td>
<td>0.8</td>
<td>0.3-1.9</td>
<td>1</td>
</tr>
<tr>
<td>MuV mumps virus</td>
<td>41 (98)</td>
<td>40 (95)</td>
<td>2</td>
<td>0.1-124</td>
<td>1</td>
</tr>
<tr>
<td>HERV human endogenous retrovirus</td>
<td>3 (7.1)</td>
<td>2 (4.8)</td>
<td>1.5</td>
<td>0.2-19</td>
<td>1</td>
</tr>
<tr>
<td>RV rotavirus</td>
<td>27 (64)</td>
<td>32 (76)</td>
<td>0.6</td>
<td>0.2-1.6</td>
<td>1</td>
</tr>
<tr>
<td>Influenza A (H1N1)</td>
<td>41 (98)</td>
<td>42 (100)</td>
<td>0</td>
<td>0.0-39</td>
<td>1</td>
</tr>
<tr>
<td>Influenza A (H3N2)</td>
<td>37 (88)</td>
<td>40 (95)</td>
<td>0.4</td>
<td>0.0-215</td>
<td>1</td>
</tr>
<tr>
<td>14 other viruses</td>
<td></td>
<td></td>
<td></td>
<td>Not Sig</td>
<td></td>
</tr>
<tr>
<td>IA-2 (positive control)</td>
<td>16 (28)</td>
<td>0 (0)</td>
<td>53</td>
<td>3.1-920</td>
<td>0.006</td>
</tr>
<tr>
<td>GST (negative control)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0</td>
<td>0-∞</td>
<td>1</td>
</tr>
</tbody>
</table>

*Bain et al. Diabetes 65:285, 2016*
T1D
Juvenile Idiopathic Arthritis

From: https://kidsgetarthritis.wordpress.com/2010/07/06
Global view of full results

2,264 results with $p_c < 10 \times 10^{-6}$ in 94 diseases and phenotypes

1,544 ChIP-seq datasets

213 diseases

- log$_{10}$ (p value)

Hemoglobin, Platelets & Red Blood Cell Traits

Autoimmune diseases + EBNA-2

Blood diseases + GATA1/TAL1 etc.
Breast Cancer

$p = 1.2 \times 10^{-46}$

<table>
<thead>
<tr>
<th>Genes</th>
<th>Tumor Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMGN1</td>
<td>T lymphocyte</td>
</tr>
<tr>
<td>MAX</td>
<td>U-87MG</td>
</tr>
<tr>
<td>ETS2</td>
<td>LoVo</td>
</tr>
<tr>
<td>GATA3</td>
<td>MCF-7</td>
</tr>
<tr>
<td>PBX1</td>
<td>MCF-7</td>
</tr>
<tr>
<td>PGR</td>
<td>T-47D</td>
</tr>
<tr>
<td>TCF12</td>
<td>RPMI-8402</td>
</tr>
<tr>
<td>ETS1</td>
<td>LoVo</td>
</tr>
<tr>
<td>BRD2</td>
<td>HEK293</td>
</tr>
<tr>
<td>FOXA1</td>
<td>ZR-75-1</td>
</tr>
<tr>
<td>POLR2A</td>
<td>HUVEC</td>
</tr>
<tr>
<td>TFAP2PC</td>
<td>MDA-MB-453</td>
</tr>
<tr>
<td>E2F1</td>
<td>MCF-7</td>
</tr>
<tr>
<td>FOXD2</td>
<td>LoVo</td>
</tr>
<tr>
<td>TCF7L2</td>
<td>MCF-7</td>
</tr>
<tr>
<td>VDR</td>
<td>THP-1</td>
</tr>
<tr>
<td>SPI1</td>
<td>OCI-Ly10</td>
</tr>
</tbody>
</table>

Coronary Heart Disease (CHD)
Best 25 TFs intersecting 55 CHD loci (April 18, 2019)
CLL & IgG glycosylation cluster with the EBNA2-related Autoimmune diseases

Chronic Lymphocytic Leukemia (CLL). Best human and EBV TF datasets intersecting 41 CLL loci (April 18, 2019)
Intersections of the 23 best TF datasets + EBV TFs with 53 Vitiligo loci (EU)
Example: CD44 locus, associated with *vitiligo*

**Informatic evaluation pipeline (MARIO)**

- dbSNP146: common
- ChIP-seq: allelic imbalance

---

**Figure Details**

- Chromosome: chr 11
- HG 19
- Position: 35293800 to 35294000
- Reference: 200 bases

**Gene Expression**

- MTA3 (GM12878)
- NFIC (GM12878)
- EBF1 (GM12878)
- BATF (GM12878)
- ATF2 (GM12878)
- EBNA-2 (Mutu)
- BCL3 (GM12878)
- ATAC (GM12878)

**SNP Information**

- rs3794102

---

*MARIO, Measurement of Allelic Ratios Informatics Operator*

At rs3794102, increased CD44 gene promotor expression using luciferase is dependent upon both the risk allele & EBV infection (T = risk and C = non-risk for vitiligo)

Conclusions

• DNA binding by TFs is concentrated in the complex genetic disease loci of many phenotypes.
• Patterns of TF association reveal TFs clustering with a subset of loci with presumed gene regulatory mechanisms.
• DNA bound by complexes containing EBNA2 are associated with genetic loci of SLE, RA, MS, T1D, IBD, CelD, & JIA.
  – Associated TFs are shared while most gene loci are not.
  – Clusters of TF binding at gene loci in these diseases are predominantly found in the Epstein-Barr virus transformed B cell, where allelic differences probably influence mechanism to alter disease risk.
  – Examples are consistent with genetic variation driving the associations.
• Epstein–Barr virus may actually cause more than MS. Our probability order is SLE, RA, T1D = IBD, CelD, & JIA.
Acknowledgements

Major contributors
• Xiaoting Chen
• Mario Pujato
• Leah Kottyan
• Matt Weirauch

Other contributors…
Daniel Miller
Avery Maddox
Carmy Forney
Albert Magnusen
Arthur Lynch
Kashish Chetal
Masashi Yukawa
Artem Barski
Nathan Salomonis

(cont.)
Kenneth Kaufman
Sreeja Parameswara
Viktoria Laurywena
Vikram Kohli
Shruti Eswar
Martha Carter

Funding (JBH & MTW)
R01 AI024717-30 Genetic Linkage in Lupus
R01 NINDS for MS
NIH R21 (NHGRI – more viruses!)
U01s (eMERGE & Autoimmune Prevention)
CCRF CpG Pilot award
CCRF Trustee Award
Lupus Research Alliance
US Dept Veterans Affairs Merit Award

Acknowledgements
• Beth Cobb
• Joshua Lee
• Connor Schroeder
• Yongbo Huang
• Xiaoming Lu
• CCHMC DNA Sequencing & Genotyping Core
• Christa Gunawan
• Ted Hong
Thank you...

The End
Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity & EBNALP and EBNA2

John B. Harley1,2,3,4,5,9*, Xiaoting Chen1,9, Mario Pujato1,9, Daniel Miller1, Avery Maddox1, Carmy Forney1, Albert F. Magnusen1, Arthur Lynch1, Kashish Chetal6, Masashi Yukawa7, Artem Barski4,7,8, Nathan Salomonis4,6, Kenneth M. Kaufman1,2,4,5, Leah C. Kottyan1,4* and Matthew T. Weirauch1,3,4,6*

Explaining the genetics of many diseases is challenging because most associations localize to incompletely characterized regulatory regions. Using new computational methods, we show that transcription factors (TFs) occupy multiple loci associated with individual complex genetic disorders. Application to 213 phenotypes and 1,544 TF binding datasets identified 2,264 relationships between hundreds of TFs and 94 phenotypes, including androgen receptor in prostate cancer and GATA3 in breast cancer. Strikingly, nearly half of systemic lupus erythematosus risk loci are occupied by the Epstein–Barr virus EBNA2 protein and many co-clustering human TFs, showing gene–environment interaction. Similar EBNA2-anchored associations exist in multiple sclerosis, rheumatoid arthritis, inflammatory bowel disease, type 1 diabetes, juvenile idiopathic arthritis and celiac disease. Instances of allele-dependent DNA binding and downstream effects on gene expression at plausibly causal variants support genetic mechanisms dependent on EBNA2. Our results nominate mechanisms that operate across risk loci within disease phenotypes, suggesting new models for disease origins.