TET2 genetic variation affects HIV viral load in ART-naïve persons

Daniel D. Murray
for the INSIGHT START study group and the FIRST study group
Background

- Compared to the general population, HIV-positive persons continue to be at greater risk of a variety of clinical events, even with optimum antiretroviral therapy (ART)
- Identifying factors that influence this risk is key for two reasons
  1. Understanding the underlying pathogenesis of HIV-disease
  2. Identification of clinically relevant biomarkers
- Previous studies, mostly in European populations, have identified SNPs in HLA and CCR5 that explain much of the genetic induced variation in HIV-VL\(^1\)\(^5\)
- Other genes are clearly involved in HIV pathogenesis and variation in these genes may also influence HIV-VL

Our targeted approach

- Can genetic information from clinical studies compliment molecular evidence and further our understanding of HIV pathogenesis?
- Target a pathway suspected to be involved in HIV-pathogenesis and determine whether variation in this pathway affects HIV-VL

**Hypothesis:** TET2 is a critical regulator of HIV-replication, and that genetic variation within this pathway will impact this function

**Aim:** Use the START and FIRST cohorts to assess the impact of genetic variation within the TET2 pathway on HIV-VL
Why TET2?

- TET2 is a host gene involved in demethylation
- TET2 involved the regulation of endogenous retroviral elements\(^1,2\)
- TET2 function has been linked to HTLV-1 (a retrovirus closely related to HIV) induced malignancy\(^3\)
- In the context of HIV, one recent study has suggested that the HIV-protein VPR selectively degrades TET2, enhancing IL-6 expression and viral replication\(^4\)
- Preliminary work in our previous GWAS\(^5\) observed a non-genome wide significant signal in TET2 that was below the minor allele frequency (MAF) cut-off used in the final manuscript, but encouraged us to explore this region further

---

The cohorts - START and FIRST

Two ART naïve cohorts from the INSIGHT network ([http://insight.ccbr.umn.edu/](http://insight.ccbr.umn.edu/))

<table>
<thead>
<tr>
<th></th>
<th>START</th>
<th>FIRST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants (genetic consent) (n = )</td>
<td>2546</td>
<td>544</td>
</tr>
<tr>
<td>Age (years) Median (IQR) or Percent</td>
<td>36 (29, 45)</td>
<td>38 (32, 44)</td>
</tr>
<tr>
<td>Female (%)</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Race (self reported)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian (%)</td>
<td>1</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Black (%)</td>
<td>23</td>
<td>57</td>
</tr>
<tr>
<td>White/other (%)</td>
<td>76</td>
<td>43</td>
</tr>
<tr>
<td>CD4+ count (cells/mm³) Median (IQR)</td>
<td>651 (585, 759)</td>
<td>220 (43, 345)</td>
</tr>
<tr>
<td>HIV RNA level (log10 copies/mL) Median (IQR)</td>
<td>4.17 (3.54, 4.66)</td>
<td>5.09 (4.53, 5.54)</td>
</tr>
<tr>
<td>Region of Residence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S. (%)</td>
<td>18</td>
<td>100</td>
</tr>
<tr>
<td>Europe/Australia/Israel (%)</td>
<td>49</td>
<td>0</td>
</tr>
<tr>
<td>South America/Mexico (%)</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>Asia² (%)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Methods - overview

Genotyping
- Custom content Affymetrix SNP CHIP – enriched for SNPs/genes involved in the immune response (including TET2)

Sample and SNP QC
- No imputation was performed

Calculation of eigenvectors
- Used in associations to control for population structure

Selection of TET2 pathway SNPs
- All SNPs (n = 888) across the TET2 pathway (IDH1 and IDH2 are regulators of TET2) with MAF > 1%

Associations with HIV-VL
- Gene and SNP level association using an additive model
Association with HIV-VL at study entry

- After QC and MAF filtering we analysed 292 and 345 SNPs for START and FIRST, respectively
- Associations with HIV-VL at study entry were performed
  1. At the gene level – using SKAT-O + gender and first four eigenvectors
  2. At the SNP level – using linear regression + gender and first four eigenvectors
- Benjamin-Hochberg procedure was used to control the false discovery rate to 5% (q-value < 0.05)
- Associations were performed independently in each cohort

Results - Associations with HIV-VL

- Gene level associations with HIV-VL in both START and FIRST
- SNP level associations in both START and FIRST
- 36 SNPs were associated with HIV-VL (q < 0.05) in one of either START or FIRST
- 15 of these SNPs were associated (q < 0.05) with HIV-VL in both cohorts
- 35/36 SNPs associated with HIV-VL were in TET2
- No gene level associations were observed in IDH1 or IDH2

Gene level associations

<table>
<thead>
<tr>
<th>Study</th>
<th>SKAT-O p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>START</td>
<td>0.000136</td>
</tr>
<tr>
<td>FIRST</td>
<td>0.000546</td>
</tr>
</tbody>
</table>

SNP level associations
Linkage Disequilibrium (LD) of SNPs associated with HIV-VL

- Two (maybe 3) groups of SNPs in LD
- One group is associated with higher HIV-VL
- One group is associated with lower HIV-VL
- All TET2 SNPs that associated with higher HIV-VL in START associated with higher HIV-VL in FIRST
- TET2 SNPs that associated with a lower HIV-VL in START associated with a lower HIV-VL in FIRST
- Strong LD structure makes identifying a causal SNP difficult
SNPs associated with HIV-VL are predominantly present in persons of Black race.

**START**
(23% Black)

- rs115930414 (n=195)

**FIRST**
(57% Black)

- rs115930414 (n=93)
- rs72963036 (n=71)

Legend:
- Blue: Black
- Red: White
- Green: Latin
- Purple: Other
Literature associations

- Most of the SNPs were part of the enrichment of the SNP CHIP and have not been reported in the literature previously
- One SNP, rs72963007, has been reported in the literature
  - Associated with an increased risk of adult T cell leukaemia caused by HTLV-1\(^1\)
  - This association was in persons of African descent
  - This SNP was present in 13% of adult T cell leukaemia (ATL) patients compared to 5% of an ethnically matched control population

Strengths and limitations of the study

Strengths

▪ Validation across independent cohorts
▪ Enrichment of TET2 in the INSIGHT CHIP
▪ Diversity of the cohort – population specific signals

Limitation

▪ Diversity of the cohort – population structure as a confounder?
▪ The mechanism of action is unclear

No genetic association model is perfect and should be viewed more as a screening tool than a final result

We need additional confirmation/validation/accumulation of supporting evidence
Conclusion

• Gene and SNP level associations indicate genetic variation within the TET2 gene affects HIV-VL.
• These data supports previous molecular evidence that TET2 is involved in HIV-replication.
• Further work is required to validate and identify the mechanism behind this change in TET2 function.
• Further work is required to identify the role of TET2 in HIV-replication.
Acknowledgements

Study participants and staff involved in the START and FIRST studies

Centre of Excellence for Health, Immunity and Infections (CHIP), Copenhagen, Denmark
- Christina Ekenberg
- Adrian Zucco
- Cameron MacPherson
- Joanne Reekie
- Marie Helleberg
- Alvaro Borges
- Jens Lundgren

University of Minnesota, USA
- Birgit Grund

Kirby Institute, Australia
- Mark Polizzotto

Tulane University Medical Center, USA
- Dahlene Fusco

APHP-Hôpital Saint Louis, Paris, France
- Julien Gras

Rigshospitalet, Copenhagen, Denmark
- Jan Gerstoft

Supported by the National Institute of Allergy and Infectious Diseases (United States), Agence Nationale de Recherches sur le SIDA et les Hépatites Virales (France), National Health and Medical Research Council (Australia), National Research Foundation (Denmark), Bundes ministerium für Bildung und Forschung (Germany), European AIDS Treatment Network, Medical Research Council (United Kingdom), National Institute for Health Research, National Health Service (United Kingdom), and the University of Minnesota. Antiretroviral drugs were donated to the central drug repository by AbbVie, Bristol-Myers Squibb, Gilead Sciences, GlaxoSmithKline/ViiV Healthcare, Janssen Scientific Affairs, and Merck.
INSIGHT Array Content – supporting slides

Modules from UK Biobank Array

<table>
<thead>
<tr>
<th>Module Name</th>
<th># Markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer’s Disease</td>
<td>803</td>
</tr>
<tr>
<td>ApoE</td>
<td>1,147</td>
</tr>
<tr>
<td>Autoimmune/Inflammatory</td>
<td>258</td>
</tr>
<tr>
<td>Blood Phenotypes</td>
<td>2,545</td>
</tr>
<tr>
<td>Cancer common variants</td>
<td>343</td>
</tr>
<tr>
<td>Cardiometabolic</td>
<td>377</td>
</tr>
<tr>
<td>HLA</td>
<td>13,519</td>
</tr>
<tr>
<td>KIR</td>
<td>1616</td>
</tr>
<tr>
<td>Lung function phenotypes</td>
<td>8,645</td>
</tr>
<tr>
<td>Common mitochondrial DNA variants</td>
<td>180</td>
</tr>
<tr>
<td>Neurological disease</td>
<td>19,791</td>
</tr>
<tr>
<td>Pharmacogenetics/ADME</td>
<td>2,856</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Module Name</th>
<th># Markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y chromosome markers</td>
<td>807</td>
</tr>
<tr>
<td>Rare variants in cancer predisposition genes</td>
<td>6,543</td>
</tr>
<tr>
<td>Rare variants in cardiac predisposition genes</td>
<td>1,710</td>
</tr>
<tr>
<td>Rare, possibly disease causing, mutations</td>
<td>13,729</td>
</tr>
<tr>
<td>eQTL</td>
<td>17,115</td>
</tr>
<tr>
<td>Fingerprint</td>
<td>262</td>
</tr>
<tr>
<td>NHGRI GWAS catalog</td>
<td>8,136</td>
</tr>
<tr>
<td>Protein truncating variants</td>
<td>30,581</td>
</tr>
<tr>
<td>Other rare coding variants</td>
<td>80,581</td>
</tr>
<tr>
<td>Genome-wide coverage for common variants</td>
<td>348,569</td>
</tr>
<tr>
<td>Genome-wide coverage for rare variants</td>
<td>37,000</td>
</tr>
</tbody>
</table>

Custom Content — 93,000 SNPs

Ddimer
Bone Mineral Density
COPD
Immune Function/Response

Hematopoiesis
Coronary Heart Disease
Pharmacogenetics/ADME
Others

725,000 unique markers represented on the array
Including an enrichment of TET2 as part of haematopoiesis