

## **Journal Club: Infektiologie und Spitalhygiene (Adrian Egli, Juli 2013)**

Science Translational Medicine, 3 July 2013 Vol 5 Issue 192 192 ra86

Karlsson EK, Harris JB, Tabrizi S, Rahman A et al.: **Natural Selection in a Bangladeshi Population from the Cholera-Endemic Ganges River Delta.**

Warum dieser Artikel:

- Genomweite Assoziationsstudie mit einem infektiologischen Fokus
- Artikel erklärt wie ein Pathogen einen Selektionsdruck auf das menschliche Genom und insbesondere das Immunsystem ausüben kann.
- Zeigt die typischen Methoden und Begriffe von GWAS in einem praktischen Kontext.

### **Einleitung:**

Cholera hat eine hohe Prävalenz in Bangladesh: 4/1000/Jahr in urbanen Slums von Dhaka (Chowdhury et al Plos Negl Trop Dis 2011).

Mit 15 Jahren haben 50% serologische Evidenz einer früheren Infektion (Glass et al JID 1985)

Cholera übt einen starken evolutionären Selektionsdruck aus, das Ganges River Delta hat die tiefste Prävalenz für die Blutgruppe O (Burua et Paguio, Ann Hum Biol 1977).

Verwandte die im selben Haushalt leben haben ein 3x höheres Risiko für eine Infektion wie unverwandte Bewohner des gleichen Haushaltes (Harris et al Plos Negl Trop Dis 2008).

Virulenz Faktoren der *Vibrio cholerae* sind gut charakterisiert: Cholera Toxin AB<sub>5</sub> -> Sekretion von Chlorid führt durch Chlorkanäle (Mekalanos et al FEMS Immuno Med Microbio 1997).

Wenig ist bekannt über Wirtsfaktoren welche zur Choleraempfindlichkeit beitragen und die Interaktion zwischen Darm und *V. cholerae*.

### **Methoden:**

Was ist eine genomweite Assoziationsstudie: Das menschliche Genom hat über 10 Millionen bekannte genetische Polymorphismen (=SNP, single nucleotide polymorphisms). Diese SNPs können mit einem Phenotyp (zB schwere Choleraerkrankung) assoziiert werden durch multivariate Regressionsanalysen.

Die verwendete Hauptmethode um genetische Faktoren zu identifizieren die mit einer erhöhten Empfindlichkeit für Cholera assoziiert sind war eine sogenannte „composite of multiple signals“ – dies ist eine statistische Methode die erlaubt eine positive Selektion von SNPs zu detektieren (Grossman et al. Science 2010).

Diese Daten werden in einer „principal component analysis“ (PCA) dargestellt. Eine PCA ist eine mathematisches Verfahren, welche eine orthogonale Transformation benutzt um ein Datensatz von Beobachtungen von möglich zusammenhängende Variablen in ein Datensatz von linearen nicht-korrelierenden Variablen zu transformieren, diese nennt man „principal components“.

Wichtige Begriffe:

Fixation Index: Populationsunterschiede erklärt durch die genetische Struktur. Werte von 0 bis 1 reflektieren die Ähnlichkeit zwischen Genen: je tiefer desto ähnlicher.

Linkage disequilibrium (LD): nicht zufällige Assoziation von Allelen an zwei oder mehr Stellen, welche von einem gleichen Chromosom ausgehen. zT findet man eine sogenannte Pyramidendarstellung.

Major Allele: Die häufigere SNP Konstellation. Im Gegensatz, minor Allele.

### **Resultate:**

Genetik der Studienpopulation (Figur 1): 42 Mutter-Vater-Kind Trios (126 Individuen) mit Bengali Ethnie aus Dhaka, Bangladesh mit einem Illumina 1M array (1,112,946 SNPs). Diese Daten wurden mit 16 anderen Populationen verglichen (International HapMap3, Singapore Genom Variation Project, Human Genome Diversity Project). Figure 1A zeigt eine klare eigenständige Population mit einer geringen Überlappung zu Indian Gujarati ( $F_{ST}=0.0047$ ). Die Population hat sich vor 52 Generationen eigenständig entwickelt (500 nach Christus).

Natürliche Selektion der Bengali Population (Figure 2): CMS kombiniert drei Indikatoren für eine positive Selektion: lange Haplotypen, hochfrequente Allele, und hoch-differenzierte Allele. Dies erlaubt die Identifikation von Regionen (oft einzelne Gene oder genomische Abschnitte) welche einen Selektionsdruck zeigen.

- Figur 2 zeigt die Signale von natürlicher Selektion in der Bengali Population im Vergleich von drei genetischen Sets.

Anreicherung von Gensets in ausgewählten genomischen Regionen in der Bengali Population: 305 ausgewählte Regionen welche zuvor identifiziert wurden, sind mit einer speziellen Software gruppiert worden. Gegenseitige vergleichende Software ermöglicht die Bestätigung von Unterschieden in der SNP Verteilung von Genen basierend auf den Populationsdaten

- Tabelle 1: MORF-IKBKG: NEMO (IKK- $\gamma$ ) Proteinbestandteil der I $\kappa$ B Kinase welche NF $\kappa$ B aktiviert. Gene für Kalium Kanäle (KCNH5 und KCNH7) und Kell Blutgruppensystem.

Testung der Topregionen zur Assoziation mit Choleraempfindlichkeit: Phentypen Vergleich von 105 Cholerapatienten und 167 unbetroffenen Individuen, die aber ein Expositionsrisiko hatten. 536 SNPs in 28 ausgewählten Regionen wurden eingeschlossen. Zusätzlich 19 SNPs welche zufällig gewählt wurden ohne vermutete Assoziation. Die Region mit der stärksten Assoziation zu Cholera war auf Chromosom 2.

- SNRNP200: small nuclear ribonucleo protein 200kDa
- CIAO1: cytosolic iron-sulfur protein assembly 1
- ITRIP1: inositol tri-phosphate receptor interacting protein-like1
- NCAPH: non-SMC condensin I complex
- TMEM127: transmembrane protein 127
- KCNH5/7: Kalium Kanäle
- RPS6KB2: ribosomal protein kinase gene
- PYCARD: Apoptosis-associated speck-like protein containing a CARD (caspase-recruitment domain)

Biologische Assoziation zwischen Genen unter natürlicher Selektion und Cholera: IKBKG, ist eine Schlüsselkomponente des NF- $\kappa$ B Signals. PYCARD (Asc<sup>-/-</sup>) knockout Mäuse und Stimulation von Macrophagen mit LPS und Cholera toxin -> weniger IL1 $\beta$  Sekretion.

#### **Diskussion:**

Interaktionsmodell: Ingenuity's IPA Software geniert eine Interaktionskarte der Bestandteile des Signalwegs der Makrophagen Stimulierung durch Cholera.

Mehrere kritische Gene welche in der Signalkaskade involviert sind, sind betroffen:

- NF $\kappa$ B Signalkaskade ist wichtig für Inflammation Aktivierung. Erhöhtes Risiko für chronisch entzündliche Darmerkrankungen bei Asiaten.
- Kaliumkanal Mutation

#### **Kritische Punkte aus meiner Sicht:**

- Offenbar gibt es SNP Unterschiede in den Populationen, ob dies aber alleine durch Cholera erklärt ist, ist fraglich; andere gram negative Bakterien, LPS zu unspezifisch.
- Gewisse Gene lassen sich nicht bestätigen in den Bestätigungskohorten.
- Für mich sind die Phentypen nicht klar definiert: „schwere“ Cholera.
- Mechanistisch lassen sich viele Dinge nicht klar erklären. Funktioneller Einfluss von SNPs auf Immunität, Zellintegrität und -funktion nicht untersucht.
- Weitere Studien wie zB Transcriptome Analysen mit RNA Sequencing erlauben klareres Bild.

#### **Take home message:**

- Genomweite Assoziationsstudien zeigen einen Zusammenhang zwischen genetischen Polymorphismen und einem Phentypen zB Choleraerkrankung.
- genaue phenotypische Beschreibungen sind nötig um eine Zuordnung zu genetischen Polymorphismen zu machen.
- Interaktionskarten der Gene zeigt neue patho-physiologische Zusammenhänge.