

## **Datenbasierte Funktionsvorhersage krankheitsrelevanter genetischer Varianten**

Matthias Arnold<sup>1,2</sup>, Johannes Raffler<sup>1</sup>, Karsten Suhre<sup>3</sup>, Gabi Kastenmüller<sup>1,4</sup>

<sup>1</sup> Institut für Bioinformatik und Systembiologie, Helmholtz Zentrum München, Neuherberg

<sup>2</sup> Department of Psychiatry and Behavioral Sciences, Duke University, NC, USA

<sup>3</sup> Department of Physiology and Biophysics, Weill Cornell Medicine-Qatar, Doha, Qatar

<sup>4</sup> Deutsches Zentrum für Diabetesforschung (DZD), Neuherberg

Genomweite Assoziationsstudien (GWAS) stellen eine effektive Methode zur Verknüpfung von genetischen Varianten (*single nucleotide polymorphisms*, SNPs) mit phänotypischen Merkmalen und Erkrankungen dar. Durch GWAS wurden mehrere zehntausend statistisch robuste Assoziationen mit hunderten unterschiedlichen Phänotypen identifiziert. Allerdings liefern diese Assoziationen in der Regel keine Hinweise darauf, welche molekularen Mechanismen den Korrelationen zwischen SNPs und assoziierten Phänotypen zugrunde liegen.

SNP-Effekte können sich auf unterschiedlichen molekularen Ebenen manifestieren, wie beispielsweise der Regulation und Expression von Genen oder der Struktur und Konzentration von Proteinen. Um über GWAS-Ergebnisse auf genetisch beeinflusste Krankheitsmechanismen rückzuschließen, die potenziell eine Translation in klinischen Anwendungen erlauben, müssen funktionelle Informationen aus unterschiedlichsten Ressourcen einbezogen werden.

Datenbanken wie *Ensembl* oder *NCBI Entrez* sammeln funktionelle Gen-, Transkript- und Protein-Annotationen auf Genomebene oder integrieren Expressionsdaten für Gene (z.B. *Gene Expression Omnibus*). Für die Interpretation von GWAS-Ergebnissen zu SNPs sind derartige Gen-zentrierte Ressourcen jedoch nur mit Einschränkungen geeignet. Zudem sind interessante Annotationen zu SNPs und Genen über verschiedene Datenbanken verteilt sowie in heterogenen Datenstrukturen abgelegt.

Wir haben daher das webbasierte Tool *SNiPA* ([www.snipa.org](http://www.snipa.org)) entwickelt, um Forscher dabei zu unterstützen, Hinweise auf die potenzielle Funktionsweise genetischer Varianten zu ermitteln. Dabei haben wir uns auf zwei Schwerpunkte fokussiert: die umfangreiche Integration bestehender Annotationsdaten für SNPs und Gene sowie intuitive und interaktive Abfrage- und Darstellungsmöglichkeiten.

Die zentrale Grundlage für *SNiPA* stellen dabei knapp 80 Millionen SNPs dar, die im Rahmen des *1000 Genomes Project* in ca. 2500 Individuen aus fünf großen Populationsgruppen nachgewiesen werden konnten. Zur Annotation dieser SNPs integrieren wir 1) Genomannotationen, also Gen-, Transkript- und Proteinsequenzannotationen und regulatorische Elemente wie Genpromotoren, Enhancerregionen, zellspezifische epigenetische Marker und transkriptspezifische microRNA-Bindestellen; 2) Effekte auf genomische Elemente, also Veränderungen der Sequenzen von Transkripten, Proteinen und regulatorischen Elementen; 3) Assoziationen aus GWAS mit molekularen und intermediären Phänotypen, also zellspezifische Genexpression (eQTLs), Proteinkonzentrationen (pQTLs) sowie Metabolitenkonzentrationen (mQTLs); 4) Assoziationen aus GWAS mit Phänotypen wie beispielsweise Körpergröße oder Nierenfunktion, aber auch klinischen Endpunkten wie Herz-Kreislauf-

Erkrankungen oder Alzheimer; 5) Scores für die vorhergesagte Konsequenz eines SNPs, die auf der Sequenzkonservierung in Säugetieren oder auf menschlichen Genomannotationen beruhen.

Die Menge an verfügbaren Annotationen, die für den einzelnen SNP integriert werden können, schwankt stark. Häufig treten SNPs jedoch nicht unabhängig voneinander, sondern regionsbasiert stark korreliert auf (*linkage disequilibrium*, LD), weshalb die Annotationen einzelner SNPs bei ausreichend ausgeprägter Korrelationsstruktur in LD-Blöcken zusammengefasst werden können. Da LD-Strukturen populationsspezifisch sind, haben wir diese für die fünf Populationsgruppen des *1000 Genomes Project* vorberechnet und in *SNiPA* integriert.

Für den Datenzugriff bietet *SNiPA* unterschiedliche Annotations- und Visualisierungsmodule. Auf der Ebene einzelner SNPs sind dies die Variantenannotation zur Abfrage aller Annotationen und Assoziationen für eine gegebene Variante sowie die Proxy-Suche, um ausgehend von einem SNP auf den zugehörigen LD-Block zurückzuschließen. Auf der variantenübergreifend Ebene liefert die Block-Annotation die kombinierte Annotation für alle Varianten für einen bestimmten Genombereich, ein bestimmtes Gen oder basierend auf einem benutzerdefinierten LD-Grenzwert. Außerdem bietet *SNiPA* einen genomweiten Variantenbrowser für eine SNP-zentrierte Sicht auf das menschliche Genom (Abb. 1), interaktive LD-Plots zur Visualisierung lokaler LD-Strukturen sowie regionale Assoziationsplots, das Locus-spezifische Pendant zum Manhattan-Plot für benutzereigene GWAS-Ergebnisse.

Alle Module bieten direkte Links zu allen Originalannotationsdatensätzen. *SNiPA* wird regelmäßig aktualisiert, um auch neue Annotationsdaten einzubinden.

#### **Literatur:**

Arnold M, Raffler J, Pfeuffer A, Suhre K und Kastenmüller G. (2015) *SNiPA*: an interactive, genetic variant-centered annotation browser. *Bioinformatics* 31(8):1334–1336

#### **Korrespondenzadresse:**

Dr. Gabi Kastenmüller  
Institut für Bioinformatik und Systembiologie  
Helmholtz Zentrum München  
Ingolstädter Landstraße 1  
85764 Neuherberg  
Tel.: +49 89 3187 3578  
g.kastenmueller@helmholtz-muenchen.de

Home

Browse

Variant Browser >

Association Maps >

Annotation

Variant Annotation >

Block Annotation >

Plots

Regional Association Plot >

Linkage Disequilibrium Plot >

Linkage Disequilibrium

Proxy Search >

Pairwise LD >

Help

Release Notes >

Documentation >

Supplemental Text >

About SNiPA >

Variant clipboard (1)

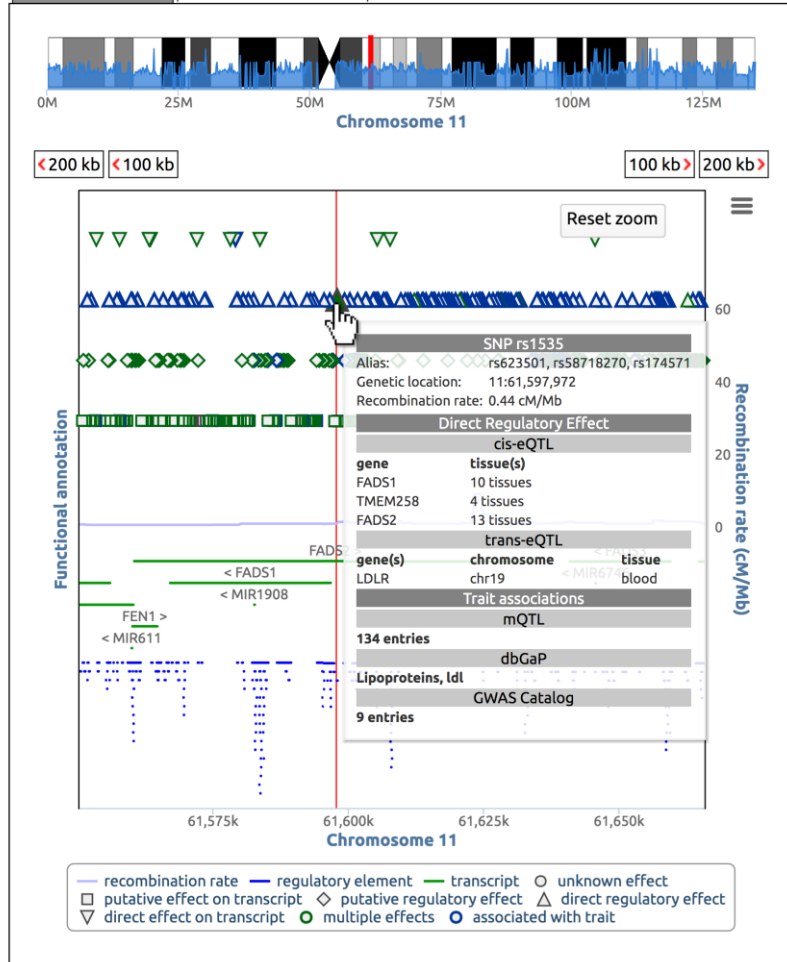
reset

rs1535 (11:61597972)

Variant Browser

Variant browser

Variant annotations



**Abb. 1:** Der SNiPA-Variantenbrowser zeigt SNPs (oben), Gene (mittig) und regulatorische Elemente (unten). Tooltips liefern kondensierte Annotationen; detaillierte Informationen sind per Mausklick verfügbar. Am Beispiel des SNPs rs1535 zeigt sich der Nutzen einer variantenzentrierten Annotation: rs1535 ist mit der Blutkonzentration verschiedener Fettstoffwechselprodukte (darunter LDL) assoziiert, außerdem mit dem Expressionslevel zweier Gene des Fettstoffwechsels (FADS1/2) sowie des Gens für den LDL-Rezeptor, einem wichtigen Regulator der Cholesterin-Homöostase. Zudem gibt es für diesen SNP eine Assoziation mit dem Ansprechen von Probanden auf Lipidsenker (Statine), welche auf die durch den LDL-Rezeptor regulierte HMG-CoA Reduktase abzielen.