Genomorganisation

Diplotyp-Architektur menschlicher Genome

MARGRET R. HOEHE

ABTEILUNG BIOINFORMATIK, MAX-PLANCK-INSTITUT FÜR MOLEKULARE GENETIK, BERLIN

10.1007/s12268-016-0741-9
© Springer-Verlag 2016

■ Hunderttausende Genome wurden bereits sequenziert, um die genetische Variabilität des Menschen zu erfassen und zu interpretieren. Dabei wurde jedoch einer grundlegenden Eigenschaft des Genoms nicht Rechnung getragen: seiner Diploidie. Menschen haben naturgemäß zwei (homologe) Chromosomensätze, je einen von Vater und Mutter. Die Anordnung der genetischen Varianten auf den beiden Homologen kann die Funktion von Genen, Phänotyp und Krankheitsentstehung entscheidend beeinflussen. Wir haben nun am Max-Planck-Institut für molekulare Genetik in Berlin eine Methode zur Haplotypsequenzierung von Genomen entwickelt (siehe dazu den Artikel auf S. 577 in dieser Ausgabe) und erstmals eine größere Anzahl diploider Genome untersucht.

Somit zu Kernfragen der Diploidie: Bei welchen Genen sind die homologen Formen gleich, bei welchen sind sie unterschiedlich? Gibt es bestimmte Klassen von Genen, die vorrangig zwei unterschiedliche homologe Formen ihrer Proteine codieren und somit ein besonderes Potenzial für funktionelle Flexibilität haben? Wie sind dabei die Mutationen zwischen den zwei Homologen verteilt? Gibt es bestimmte Verteilungsmuster? In einer erstmaligen Analyse diploider Genome auf Populationsebene [1], die 14 molekular und 372 statistisch haplotypisierte Genome europäischen Ursprungs miteinbezog, konnten wir bestimmte Charakteristika der Architektur diploider Gene und Genome herausarbeiten.

Proteindiplotypen sind signifikant häufiger in ca. 4.000 Genen

Bei der Analyse der insgesamt 386 Genome fiel auf, dass der Anteil der Gene, die zwei unterschiedliche Proteinformen codieren, in jedem Genom etwa gleich groß war: um die

20 Prozent. Handelt es sich dabei jeweils um eine zufällige Auswahl von Genen, oder existiert ein gemeinsamer Satz von Genen, die häufiger als Diplotypen vorkommen? In der Tat konnten wir 4.269 Gene extrahieren [1], die signifikant häufiger „Proteindiplotypen“ codieren, wodurch Genfunktionen differenziell ausgeübt oder diversifiziert werden können. Wir haben diesen Gensatz auch als *common diplotypic proteome* (CDP) bezeichnet. Simulationsstudien bestätigten, dass das CDP nicht durch zufällige Verteilung von Mutationen auf die autosomalen Gene hätte zustande kommen können. Das CDP könnte ein wichtiges „modulierendes Prinzip“ darstellen, indem es eine große Vielfalt zellulärer und physiologischer Reaktionen in individuellen Organismen bewirkt – die Grundlage für adaptive Prozesse und phänotypische Vielfalt. Das CDP war hochsignifikant angereichert mit funktionellen Genklassen, die eine wichtige Rolle bei der rezeptorvermittelten Signaltransduktion, Zell-Zell-Kommunikation, Transkriptionsregulation und bei Immunprozessen spielen [1]. Dies weist auf eine besondere Bedeutung der Diploidie für die Aufrechterhaltung der funktionellen Flexibilität dieser Funktionen hin. Sind die „falschen“ Proteinformen aktiv, könnte das CDP den Boden für die Entwicklung von Krankheiten bereiten, was durch Anreicherung einer Vielzahl von Krankheitsgenen gestützt wird [1].

Cis-Konfigurationen sind signifikant häufiger als trans-Konfigurationen

Cis-Konfigurationen von Mutationen, die eine Gen- bzw. Proteinform intakt lassen (siehe S. 577), sollten häufiger als trans-Konfigurationen im menschlichen Genom vorkommen, um eine intakte Funktion des Organismus zu gewährleisten. So bestimmten wir zunächst

für jedes einzelne Genom die jeweilige Anzahl der Gene, die *cis*- bzw. *trans*-Konfigurationen von potenziell funktionsverändernden Mutationen hatten. Ohne jede Ausnahme kamen in jedem einzelnen der 14 molekular und 372 statistisch haplotypisierten Genome *cis*-Konfigurationen signifikant häufiger als *trans*-Konfigurationen vor, bei einem durchschnittlichen *cis/trans*-Verhältnis von ungefähr 60:40. Nach den genomweiten Analysen nun die Einzelgen-Betrachtung: Für jedes Gen wurden die in der jeweiligen Stichprobe vorkommenden *cis*- und *trans*-Konfigurationen gezählt und ein genbasierter *cis/trans*-Quotient bestimmt. Dabei ließ sich eine große Gruppe von mutierten autosomalen Genen identifizieren, die signifikant häufiger *cis*-Konfigurationen von potenziell funktionsverändernden Mutationen zeigten (*cis-abundant genes*), und eine kleinere Gruppe mit signifikant häufigeren *trans*-Konfigurationen (*trans-abundant genes*). Diese beiden Gruppen von Genen konnten weiter funktionell differenziert werden [1]. Darüber hinaus waren vorrangig „*trans*-konfigurierte“ Gene signifikant häufiger mit Krankheiten assoziiert.

Die Ergebnisse zur Architektur diploider Gene und Genome zeigen erstens, dass Mutationen nicht zufällig auf autosomalen Genen verteilt sind, sondern vorrangig in bestimmten Funktionsklassen vorkommen, deren potenzielle funktionelle Flexibilität sie somit konstituieren, und zweitens, dass, wo sie vorkommen, die Mutationen vorrangig auf demselben Homolog angeordnet sind, bei einem *cis/trans*-Verhältnis von etwa 60:40 ausnahmslos in jedem Individuum. Drittens zeigen sie, dass *cis-abundant genes* und *trans-abundant genes* existieren. Haplotyp-/Diplotyp-Analysen sind folglich unverzichtbar für ein tieferes Verständnis der genetischen Variabilität des Menschen und seiner Biologie und damit für eine personalisierte Genomik und Medizin. ■

Literatur

[1] Hoehe MR, Church GM, Lehrach H et al. (2014) Multiple haplotype-resolved genomes reveal population patterns of gene and protein diploypes. Nat Commun 5:5569

Korrespondenzadresse:

Dr. med. Dipl. Psych. Margret R. Hoehe
AG Diploid Genomics
Abteilung Bioinformatik
Max-Planck-Institut für molekulare Genetik
Ihnestraße 63-73
D-14195 Berlin
Tel.: 030-8413-1614
Fax: 030-8413-1152
hoehe@molgen.mpg.de
www.molgen.mpg.de/157216/Diploid-Genomics