

Paper of the Quarter – QII/2023 – [mitoNET](#)

Aberrant splicing prediction across human tissues

Nat Genet. 2023 May. [->PubMed-Link<<](#)

Nils Wagner, Muhammed H Çelik, Florian R Hölzlwimmer, Christian Mertes, Holger Prokisch, Vicente A Yépez, Julien Gagneur

Genetische Varianten, d.h. Mensch-zu-Mensch-Variationen des Erbguts, können drastische Auswirkungen auf zelluläre Prozesse haben. Einige von ihnen können schädlich sein und Krebs oder seltene Krankheiten verursachen. Aber welche? Unsere Genome tragen viele Tausende von seltenen – oft einzigartigen – genetischen Varianten, die harmlos sind. Das Aufspüren von Varianten mit drastischen Auswirkungen ist ein Problem wie die Nadel im Heuhaufen zu finden. Das Problem wird zusätzlich dadurch erschwert, dass einige Varianten ihre Auswirkungen nur in bestimmten Geweben wie dem Gehirn oder dem Herzen entfalten.

Hier haben wir einen Benchmark-Datensatz erstellt, der genetische Varianten mit gewebespezifischem aberrantem Spleißen verknüpft, einem Prozess, der den Sequenzinhalt von RNAs drastisch verändert, indem ganze Segmente von vielen, oft Hunderten von Nukleotiden verloren gehen oder hinzugefügt werden. Fehlerhaftes Spleißen führt häufig zu nicht funktionierenden RNAs, was zu Krankheiten führen kann.

Anhand dieses einzigartigen Datensatzes haben wir herausgefunden, dass etablierte KI-Modelle, die das Spleißen anhand der DNA-Sequenz modellieren, bei der Vorhersage von gewebespezifischem, fehlerhaftem Spleißen bemerkenswert ungenau sind. Es ist uns gelungen, die Präzision dieser Modelle drastisch zu verbessern, indem wir kartierten, wo im Genom das Spleißen stattfindet, in welchen Geweben und mit welchen Raten, selbst bei schwachem Spleißen, das nur ~1 % der RNA-Moleküle betrifft. Durch die Integration dieser Daten in ein biophysikalisches Spleißmodell konnte die Präzision der KI-Modelle um das Dreifache gesteigert werden.

In einem zweiten Schritt haben wir gezeigt, wie man nicht nur die DNA der Patienten, sondern auch RNA aus zugänglichen Geweben, wie Blut oder Haut, nutzen kann, um fehlerhaftes Spleißen in nicht zugänglichen Geweben, wie dem Gehirn oder dem Herzen vorherzusagen. Dies ist von großer Bedeutung, da viele seltene Krankheiten Organe betreffen, bei denen eine Probenentnahme nur durch besonders invasive Eingriffe möglich ist. Das Modell setzt ein einfaches Prinzip um: Wenn wir es nicht vorhersagen können, messen wir es einfach! Das Modell nutzt direkte Messungen von fehlerhaftem Spleißen in zugänglichen Geweben für diejenigen Gene, die sich auf die gleiche Weise verhalten wie in den Geweben von Interesse, und verlässt sich ansonsten auf DNA-basierte Vorhersagen. Dies steigerte die Leistung noch einmal um das Zweifache.

Wir verwenden das von uns entwickelte Modell AbSplice jetzt, um innerhalb europäischer Konsortien die genetische Ursache von Patienten mit seltenen Krankheiten zu ermitteln. Außerdem verwenden wir es, um verbesserte genetische Risikobewertungen für häufige Krankheiten mit allen genetischen Varianten zu entwickeln, seien sie selten oder häufig, die das einzigartige genetische Make-up von jedem von uns ausmachen.

Autor*innen: Nils Wagner, Muhammed H Çelik, Florian R Hölzlwimmer, Christian Mertes, Holger Prokisch, Vicente A Yépez, Julien Gagneur

Kontakt: wagnern@in.tum.de, gagneur@in.tum.de

Paper of the Quarter – QII/2023 – [mitoNET](#)

Aberrant splicing prediction across human tissues

Nat Genet. 2023 May. [->PubMed-Link<<](#)

Nils Wagner, Muhammed H Çelik, Florian R Hölzlwimmer, Christian Mertes, Holger Prokisch, Vicente A Yépez, Julien Gagneur

Genetic variants, i.e. person-to-person variations in the genome, can have drastic effects on cellular processes. Some of them can be harmful, causing cancer or rare diseases. But which ones? Our genomes carry many thousands of rare – often unique – genetic variants which are harmless. Pinpointing variants with drastic effects is a needle in a haystack problem. The problem is even harder as some variants exert their effects only in specific tissues such as the brain or heart.

Here we have established a benchmark dataset linking genetic variants to tissue-specific aberrant splicing, a process that drastically alters the sequence content of RNAs by losing or adding entire segments of many, often hundreds, of nucleotides. Aberrant splicing often results in non-functional RNAs which can lead to disease.

With this unique dataset at hand, we found that established AI models, which simulate splicing from DNA sequence, are remarkably imprecise in predicting tissue-specific aberrant splicing. We managed to drastically improve the precision of these models by mapping where in the genome splicing is taking place, in which tissues, and at what rates, even for weak splicing, which only affects ~1% of RNA molecules. Integrating this data with a biophysical model of splicing, boosted up the AI model precision by 3-fold.

In a second step, we showed how to leverage not only DNA of the patients, but also RNA obtained from accessible tissues, say blood or skin, to predict aberrant splicing in non-accessible tissues such as the brain or heart. This is important, as many rare diseases affect organs from which samples can only be obtained through particularly invasive procedures. The model implements a simple principle: if we cannot predict, let's just measure it! The model leverages direct observations of aberrant splicing in accessible tissues for those genes behaving the same way than in the tissues of interest, and rely on DNA-based predictions otherwise. These boosted the performance by another 2-fold.

We are now using our model AbSplice to help pinpointing the genetic cause of rare disease patients within European consortia. Furthermore we use it to develop improved genetic risk scores for common diseases using all genetic variants, be they rare or frequent, that constitute the unique genetic make-up of every one of us.

Author(s): Nils Wagner, Muhammed H Çelik, Florian R Hölzlwimmer, Christian Mertes, Holger Prokisch, Vicente A Yépez, Julien Gagneur

Contact: wagnern@in.tum.de, gagneur@in.tum.de