SNPs in NAGNAG acceptors are highly predictive for variations of alternative splicing and potentially relevant for human diseases

Michael Hiller[1], Klaus Huse[2], Karol Szafranski[2], Niels Jahn[2], Jochen Hampe[3], Stefan Schreiber[3], Rolf Backofen[1] and Matthias Platzer[2]

[1] Institute of Computer Science, Chair for Bioinformatics
Albert-Ludwigs-University Freiburg,
Georges-Koehler-Allee 106, 79110 Freiburg, Germany
email: hiller@informatik.uni-freiburg.de

[2] Genome Analysis,
Leibniz Institute for Age Research - Fritz Lipmann Institute,
Beutenbergstr. 11, 07745 Jena, Germany
email: mplatzer@fli-leibniz.de

[3] Institute of Clinical Molecular Biology
Christian-Albrechts-University Kiel
Schittenhelmstr. 12, 24105 Kiel, Germany
email: s.schreiber@mucosa.de

Introduction

Much interest focuses on single nucleotide polymorphisms (SNPs) that are located in coding regions since they may alter the protein sequence. However, SNPs can also influence constitutive and alternative splicing by affecting donor and acceptor splice sites, branch points, exonic and intronic splicing enhancers and silencers. Recently, splicing mutations have been suspected to be the most frequent cause of hereditary diseases and an increasing number of SNPs has been described that cause diseases by a change or disruption of the normal splicing pattern.

Previously, we reported the widespread occurrence of subtle alternative splice events that insert or delete the sequence NAG (N = A,C,G, or T) in mRNAs [1]. This happens if both AGs of a NAGNAG or tandem acceptor can be chosen by the spliceosome. In these structures, we denote the upstream AG as the ”E acceptor” giving rise to the ”E transcript”, since part of the tandem will be exonic while the whole tandem is intronic for the ”I acceptor” (Figure 1).

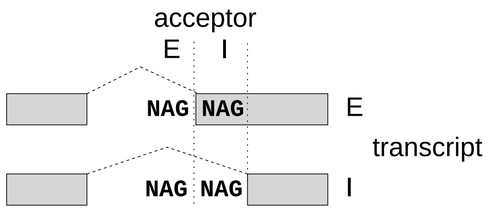

Figure 1: Nomenclature. E: 3’ half of the NAGNAG motif becomes part of the exon; I: the NAGNAG motif is completely retained in the intron.

Based on the NAGNAG motif, **31%** of the HAGHAG acceptors (H = A,C, or T) but only only **1.7%** of the remaining NAGNAGs are alternatively spliced (as derived from searching EST and mRNA databases). Thus, we propose to subdivide all NAGNAG acceptors into plausible (HAGHAG) and implausible (GAGHAG, HAGGAG, or GAGGAG) acceptors.

Identification of SNPs affecting NAGNAG acceptors

- We selected all SNPs from dbSNP that affect a NAGNAG acceptor [2]
- With respect to the human reference genome sequence we classified those SNPs into (Figure 2)
  (A) destroy a NAGNAG acceptor
  (B) create a NAGNAG acceptor
  (A) change an ’N’ in the NAGNAG motif

⇒ **total of 121 SNPs**

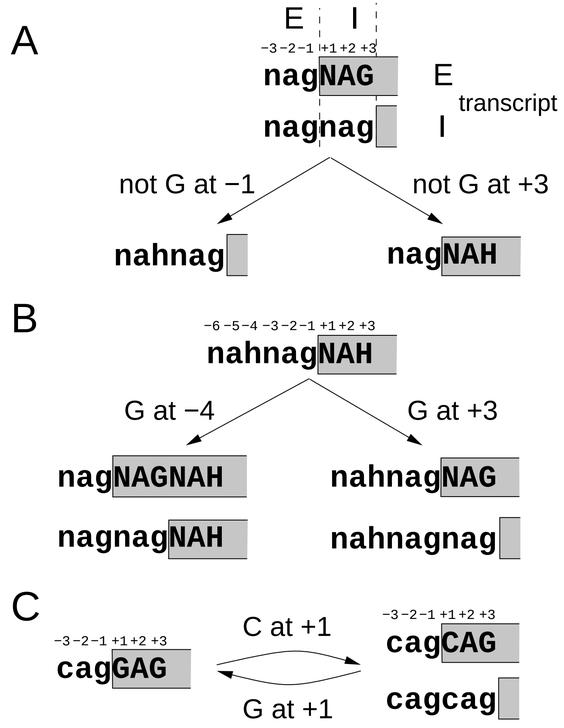

Figure 2: (A) SNP alleles at position -2, -1, +2, or +3 of a NAGNAG acceptor destroy this motif by affecting the E (left) or I (right) acceptor, thus preventing alternative splicing. (B), SNP alleles at intron positions -5 and -4 can create a novel E acceptor (left) and, at exon positions +2 and +3, a novel I acceptor (right), thus yielding a NAGNAG motif that may allow alternative splicing. (C) SNP alleles at position -3 or +1 of a NAGNAG acceptor can convert a plausible NAGNAG that allows alternative splicing (left) to an implausible one that allows only the expression of one transcript (right), or vice versa.

Identification of splice-relevant NAGNAG SNPs

Not all of the 121 SNPs are likely to influence alternative splicing at NAGNAG acceptors:

- 57 of the 121 are presumably **not relevant** for NAGNAG splicing since they
  - create an implausible NAGNAG
  rs1132591 TAG|GAT and TAG|GAG ⟹ TAGGAG is unlikely to be alternatively spliced
  - destroy an implausible NAGNAG
  rs2292402 GAGCAG| and GTGCAG| ⟹ only the CAG should function as an acceptor
  - change an implausible NAGNAG to another implausible one
  rs9866111 TAG|GAG and CAG|GAG ⟹ both alleles are unlikely to be alternatively spliced
  - change an plausible NAGNAG to another plausible one
  rs4149853 CAGCAG| and TAGCAG| ⟹ both alleles are very likely to be alternatively spliced

- 64 of the 121 are **relevant** for NAGNAG splicing since they
  - create a plausible NAGNAG
  rs1127307 TAG|CGG and TAG|CAG ⟹ TAGCAG is confirmed by 109:7 E:I transcripts
  - destroy a plausible NAGNAG
  rs2298847 TAGCAG| and TTGCAG| ⟹ TAGCAG is confirmed by 55:21 E:I transcripts
  - change a plausible NAGNAG to an implausible one or vice versa
  rs8176139 TAG|CAG and GAG|CAG ⟹ TAGCAG is confirmed by 27:7 E:I transcripts

- **18 of the 64 (28%) affected NAGNAG acceptors have EST/mRNA evidence for alternative splicing**

⇒ **consider 64 SNPs as relevant for alternative splicing at NAGNAG acceptors**

NAGNAG motif is necessary and sufficient for alternative splicing

SNPs that create/destroy a NAGNAG acceptor represent ”knockout experiments made by nature”. We used these SNPs to investigate the assumed correlation between NAGNAG acceptor genotypes and the appearance of E and I transcripts. We selected four SNPs that affect EST-confirmed HAGHAG acceptors and observed E and I transcripts in cells with at least one HAGHAG allele, whereas cells that do not have a HAGHAG acceptor allele produced only one transcript (Figure 3). This strict correlation between NAGNAG alleles and alternative splicing confirm that NAGNAG motifs are necessary for this type of alternative splicing.

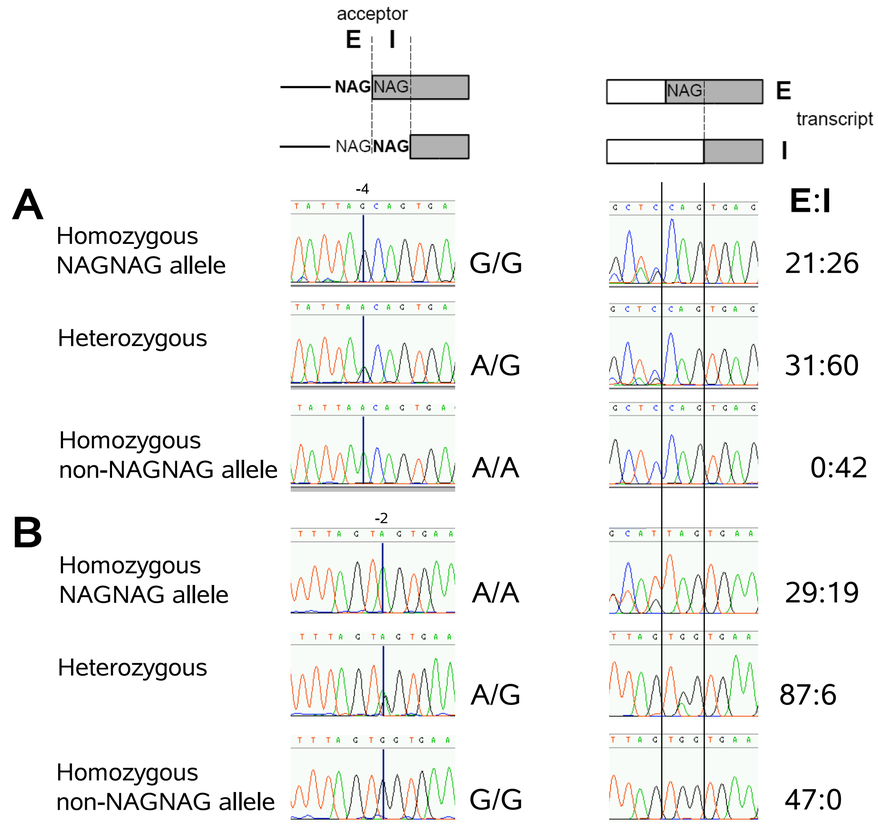

Figure 3: (A) SNP rs2245425 affecting the E acceptor of *TOR1AIP1* exon 3 leads to the exclusive expression of the I transcript from the A allele. (B) SNP rs2275992 affecting the I acceptor of *ZFP91* exon 5 leads to the exclusive expression of the E transcript from the G allele. E:I transcript ratio was determined by counting subcloned and sequenced RT-PCR fragments.

Next, we asked whether a single mutation that changes a non-NAGNAG to a NAGNAG acceptor is sufficient to enable alternative NAGNAG splicing. We used the chimpanzee genome sequence to determine the non-ancestral allele variant of these 64 SNPs. In 43 cases, the plausible NAGNAG is non-ancestral and therefore gained in the recent human evolution. Consistent with our assumption, we found **EST evidence of alternative NAGNAG splicing for 7 of these 43 (16%) NAGNAGs**. Furthermore, we selected a non-ancestral plausible NAGNAG without EST evidence for experimental investigation. As expected, in leukocytes of individuals heterozygous or homozygous for the NAGNAG allele of rs5248 (CAACAG → CAGCAG), we observed E and I transcripts in the ratios 4:14 and 11:7, respectively. These findings suggest that NAGNAG motifs are sufficient for alternative splicing in the context of a previously non-NAGNAG acceptor.

⇒ **SNPs in NAGNAGs are highly predictive for variations of alternative splicing**

Three protein isoforms by non-silent SNPs

- 15 of the 64 (23%) SNPs are translationally non-silent
- They change the I acceptor and the amino acid sequence of the E protein (Figure 4)

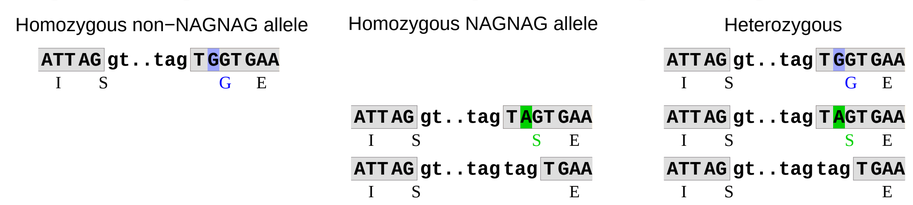

Figure 4: The non-silent SNP rs2275992 in *ZFP91*. Homozygotes express either one or two isoforms. Heterozygosity results in three different proteins.

⇒ **non-silent SNPs result in three different protein variants**

Disease relevance of NAGNAG SNPs

The disease relevance of a NAGNAG SNP was demonstrated by Maugeri et al. [3] for the *ABCA4* gene where a mutation (2588G→C) changes the acceptor TAGGAG→TAGCAG and leads to the expression of E and I transcripts. Our study exactly predicts this mutation outcome. This mutation has a much higher frequency in patients with **Stargardt disease 1** (STGD1) and is assumed to be a mild mutation that causes STGD1 in combination with a severe *ABCA4* mutation.

Further examples from dataset of 64 NAGNAG SNPs:

- rs9644946 changes the exon 8 acceptor of *GOLGA1* from AAATAG to AAGTAG and would lead to the insertion of an inframe TAG for the E transcript and potentially nonsense-mediated mRNA decay. *GOLGA1* codes for an autoantigen associated with **Sjogren syndrome**.
- rs363209 changes the exon 7 acceptor of *APPBP1* from AAACAG to AAGCAG. The APP-BP1 protein binds APP and interacts with UBE1C in the process of neddylation. APP plays a central role in **Alzheimer disease** and **Down syndrome**.
- other examples include **Breast cancer** (*BRCA1*), **Down syndrome** (*GART*), and **Asthma** (*IL19*)

⇒ **18 of the 64 (28%) SNPs occur in known disease genes**
⇒ **candidates for functional analysis and association studies**

[1] Hiller et al. ”Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity.” *Nature Genetics*, 2004. **36**(12):1255-1257.

[2] Hiller et al. ”Single-nucleotide polymorphisms in NAGNAG acceptors are highly predictive for variations of alternative splicing.” *Am. J. Hum. Genet.*, 2006. **78**(2):291-302.

[3] Maugeri et al.”The 2588G→C mutation in the ABCR gene is a mild frequent founder mutation in the Western European population and allows the classification of ABCR mutations in patients with Stargardt disease.” *Am. J. Hum. Genet.*, 1999. **64**(4):1024-35.