LETTER TO THE EDITOR

GATA1 Mutations in Myeloproliferative Disorders: Nomenclature Standardization and Review of the Literature

Alessandra Splendore,1* Isis Q. Magalhães,2 and Maria S. Pombo-de-Oliveira3

1Divisão de Genética, Instituto Nacional de Câncer, Rio de Janeiro, Brasil; 2Departamento de Hematologia/Oncologia Pediátrica SES-DF, Brasília, Brasil; 3Divisão de Medicina Experimental, Instituto Nacional de Câncer, Rio de Janeiro, Brasil

Communicated by Stylianos Antonarakis

The GATA1 gene (MIM #305371), which maps to Xp11.23, is a member of the GATA binding family of transcription factors that are specifically expressed in erythrocytes, mast cells, megakaryocytes, and eosinophil progenitors [Fujisawa et al., 1996]. GATA1 expression normally suppresses the proliferation of erythrocytes and megakaryocytes precursors while promoting their differentiation [Izraeli, 2004]. Since males are hemizygous for GATA1, and one copy of this gene is silenced by X chromosome inactivation in females, only one allele is expressed per cell. GATA1 consists of six exons extending along 7 kb and contains an ORF of 1239 nucleotides starting in exon 2. The 413-amino acid GATA-1 protein has an N-terminal transactivation domain and two zinc-finger domains. Experimental data show that GATA-1 binds to DNA and a variety of cofactors, including friend of GATA-1 (FOG, encoded by the ZFPFM1 gene, MIM# 601950), and the runt-related transcription factor 1 (RUNX1, MIM# 151385). GATA1 also encodes an alternative 330-amino-acid-long isoform named GATA-1 s. This shorter isoform, which is translated from a ATG codon located in exon 3 (Met84), lacks the transactivation domain but retains the DNA-binding ability and interacts normally with FOG [Calligaris et al., 1995; W echsler et al., 2002]. The proposed mechanisms for the production of GATA-1 s include alternative transcription initiation (using the Met84 as initiation codon) and alternative splicing of exon 2 [Calligaris et al., 1995; Rainis et al., 2003]. Although both GATA-1 isoforms are found in mouse embryonic tissue, their relative proportion varies during development, which suggests that the transcriptional activity of GATA-1 may be modulated by the relative ratios of the two isoforms [Calligaris et al., 1995].

It is a well known fact that children with trisomy 21, or Down syndrome (DS, MIM# 190685), have a higher incidence of acute leukemia, and their risk for acute megakaryoblastic leukemia (AMKL) in particular is increased 500-fold in comparison with normal children [Ross et al., 2004]. Wechsler et al. [2002] reported that somatic mutations in GATA1 were found in the proliferating clones of DS children with AMKL. These results were later confirmed by other authors and expanded to show that the self-regressing leukemia that affects up to 20% of DS neonates (transient myeloproliferative disorder (TMD)) was also characterized by GATA1 somatic mutations [reviewed in Crispino, 2004]. Mutations in GATA1 were found in 73 of 78 (93%) DS-TMD patients, and in 70 of 83 (84%) cases of DS-AMKL [Wechsler et al., 2002; Groet et al., 2003; Hitzler et al., 2003; Mundschau et al., 2003; Xu et al., 2003; Rainis et al., 2003; Hirose et al., 2003; Ahmed et al., 2004; McElwaine et al., 2004]. GATA1 mutations can be identified by sequencing either genomic DNA or cDNA. The genomic approach of exon-by-exon PCR amplification can also be applied to archival samples, including autopsy material, neonatal blood spots, and bone marrow smear slides [Taub et al., 2004; Ahmed et al., 2004; Magalhães et al., in press]. On the other hand, cDNA-based screening is more sensitive because deletions that encompass the whole exon 2 may be missed in amplifications using intronic primers [Xu et al., 2003]. The shortcoming of using only cDNA is that once a product of alternative splicing of exon 2 (joining exon 1 to exon 3) is identified, we cannot determine whether the underlying mutation is a large deletion or a single-base change at the intronic splice site. Although there are no pseudogenes or homologous sequences that could complicate mutation detection in this gene, tumor mass affects the sensitivity of the assays. If the blast proportion in the sample is small, direct sequencing or cloning may fail to identify mutant cells. In such cases, denaturing acrylamide gel electrophoresis with either radioactive labeling or silver staining has been shown to be useful for detecting low-copy mutant clones [Ahmed et al., 2004; Magalhães et al., in press].

Over 100 different somatic mutations have been found in GATA1, most of which (70%) are frameshift-causing small deletions, insertions, duplications, and indels that occur prior to the codon for Met84. Other mutations include changes at the splice donor site of exon 2, missense alterations in the initiation codon (Met1), premature stop codons within exon 2, or large deletions involving the whole or a large part of exon 2. The majority of GATA1 mutations are located in exon 2, but two patients in a study by Groet et al. [2003] had a 2-pb duplication (c.231_232dupGT) in exon 3. Since exon 3 contains only 29 nucleotides upstream of the Met84 codon, the probability of a TMD/AMKL-related mutation occurring in this exon is small. Nevertheless, exon 3 should also be included in screening

Received 21 March 2005; accepted manuscript 8 June 2005.

*Correspondence to: Alessandra Splendore, Ph.D., Centro de Pesquisa-Instituto Nacional de Câncer, Divisão de Genética, Rua André Cavalcanti, 37-4' Andar, CEP 20231-050 Rio de Janeiro, RJ, Brasil. E-mail: ale.splendore@terra.com.br

DOI: 10.1002/humu.20233

Published online in Wiley InterScience (www.interscience.wiley.com).
protocols, especially when screening of exon 2 fails to detect a mutation. All mutations are predicted to disrupt the translation of GATA-1 while preserving that of GATA-1s.

When we try to compare mutations described by different authors, we notice that three different criteria for mutation nomenclature have been used. Although mutations were described according to the same reference sequence (NM_002049), the numbering of nucleotides was not consistent among different authors. In some reports, nucleotide +1 is assigned to the first nucleotide of the reference sequence, including the 5' untranslated region (UTR) of exons 1 and 2 [Groet et al., 2003; Hitzler et al., 2003; Rainis et al., 2003; Ahmed et al., 2004; Taub et al., 2004; McElwaine et al., 2004; Carpenter et al., 2005]. In two reports [Xu et al., 2003; Hirose et al., 2003], +1 was assigned to the first nucleotide of exon 2 (number 94 in the reference sequence), including 19 bases of the 5' UTR prior the beginning of the coding region. Other authors [Wetscher et al., 2002; Mundschau et al., 2003; Magalhães et al., 2005] assigned +1 to the first adenine of the initiation codon (number 113 in the reference sequence), in accordance with HGVS mutation nomenclature guidelines (www.hgvs.org/mutnomen/).

In addition to discordant numbering, several reports failed to discriminate whether an inserted sequence was in fact a duplication (especially if the case involved a single-base duplication). Moreover, other studies provided only the number of inserted nucleotides, without specifying which nucleotides were inserted. In the postgenomic era, it is important to consistently use a standard nomenclature for mutations to facilitate comparisons of different reports, highlight recurring mutations, and prevent duplication of data [den Dunnen & Paalman, 2003]. The online Table S1 (available online at www.interscience.wiley.com/pages/1059-7794/suppmat) presents all GATA1 mutations reported to date that are associated with myeloproliferative disorders, as originally described and renamed to conform with current HGVS mutation nomenclature guidelines.

The differences in nomenclature obscured the fact that in some instances, more than one author independently reported the same mutation (Table 1). Since all of the mutations discussed herein are acquired somatically, most of them are private. The presence of recurring mutations may indicate a mutation-prone sequence context; however, a search for repetitive sequence motifs in exon 2 conducted using OligoRep (http://wwwmgs.bionet.nsc.ru/mgs/programs/oligorep/) failed to correlate the presence of direct, inverted, symmetric, or complementary repeats with the sites of any of the recurring mutations.

Ahmed et al. [2004] described four patients with different GATA1 mutations in multiple clones, indicating that GATA1 mutations show a high frequency in DS neonates. One likely explanation for this finding is that an extra chromosome 21 somehow predisposes an individual to somatic mutations. However, this scenario seems unlikely because DS individuals are not predisposed to a variety of different types of cancer [Ahmed et al., 2004]. An alternative explanation is that this gene has a high rate of somatic mutations in hematopoietic cells. We speculate that while somatic mutations occur along the whole gene, only those that preserve the expression of GATA-1s are selected, and those that disrupt both isoforms either do not confer a proliferative advantage or are detrimental to the survival of the clone. While this high mutation rate may occur not only in DS children but also in the general pediatric population, only in the presence of an extra chromosome 21 will such mutations confer a proliferative advantage, resulting in TMD [Crispino, 2004].

Reinforcing this notion is the observation that TMD occurs almost exclusively in association with DS. When TMD is reported without a DS phenotype, patients are mosaic for trisomy 21 or have a somatically acquired trisomy 21 restricted to the proliferative clone [reviewed in Sandoval et al., 2004; Magalhães et al., 2005; Carpenter et al., 2005]. Conversely, one family with an inherited c.220G>C mutation, in the absence of trisomy 21, presented with anemia and neutropenia, but not leukemia. These findings highlight the fact that cooperation between GATA-1s and a dosage effect of genes on chromosome 21 is required to promote abnormal proliferation of megakaryoblasts. The main candidate for this mechanism is the megakaryocyte-specific transcription factor RUNX1, located on 21q22.3, which interacts with GATA-1 but not with GATA-1s [Hitzler and Zipursky, 2005]. There is an ongoing search for other genetic factors that might differentiate the self-regressing TMD from AMKL.

### References


