© American College of Medical Genetics and Genomics

 Genetics inMedicine

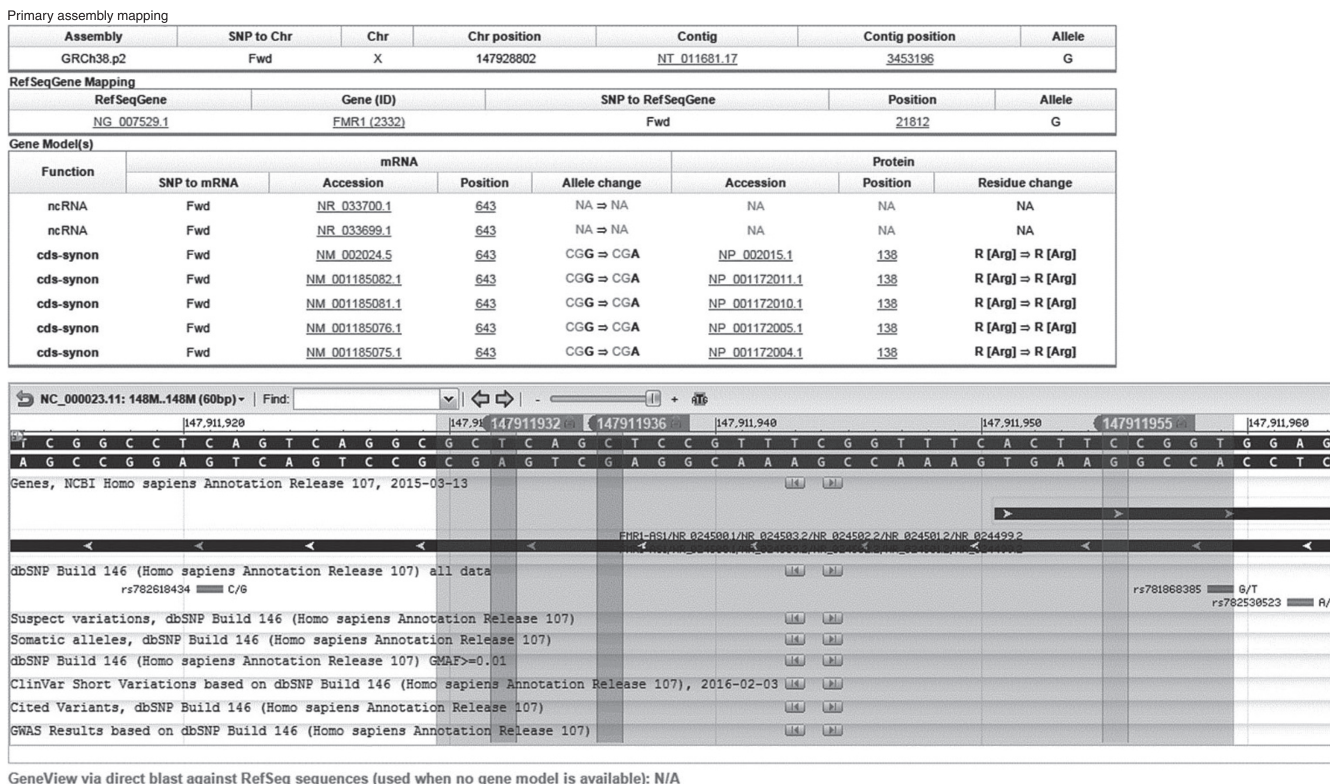
Use of bioinformatic tools in primer validation

To the Editor: We read with great interest the article "Similar Prevalence of Expanded CGG Repeat Lengths in the Fragile X Mental Retardation I Gene Among Infertile Women and Among Women With Proven Fertility: A Prospective Study," by De Geyter et al.,[1] which describes an interesting approach to measuring the CGG repetitions in the 5′ untranslated region (5′ UTR) on the *FMR1* gene. However, as we were validating the approach, we found some issues regarding the first set of locus-specific primers for the conventional polymerase chain reaction (PCR), the first step of its methodology.

We performed an initial analysis using primer 3 with the sequence of the 5′ UTR region containing the CGG repetition site (GenBank accession number L29074.1) and the primers reported by De Geyter et al.[1] However, to our surprise, the output showed no match for any of the primers reported. Intrigued, we performed a BLAST, and both primers (forward and reverse) showed 100% identity with the human FMR1 gene. To clarify this, we decided to search the primer's sequence using the National Center for Biotechnology Information's Variation Viewer. Interestingly, both of the primer's sequences presented a match only for the *FMR1* gene, although in an unexpected manner. The sequence presented as the forward primer was in the negative strand, in 3′–5′ sense, making it, by definition, the reverse primer (R′). Likewise, the sequence presented as the reverse primer was in the positive strand, in 5′–3′ sense (F′). Moreover, both of the sequences had reported variants according to the Variation Viewer. The true reverse primer (R′) had two variants reported in the middle of the target sequence, and the true forward primer (F′) had one variant reported at the 3′ end (**Figure 1**).

When we repeated the analysis using the new primer's positions with primer 3 (ref. 2), the R′ primer resulted in no match, a fact that we attribute to the reported variants. The F′ primer validation, although a match, showed high Tm (71.88°C) and high hairpin stability.

The primers reported by De Geyter et al.,[1] although corresponding to the desired region, were not considered optimal when analyzed by different bioinformatics tools. First, variants found in the target region in both primers could lead to mismatches between targets and primers in the samples. Even



Primary assembly mapping

Assembly	SNP to Chr	Chr	Chr position	Contig	Contig position	Allele
GRCh38.p2	Fwd	X	147928802	NT_011681.17	3453196	G

RefSeqGene Mapping

RefSeqGene	Gene (ID)	SNP to RefSeqGene	Position	Allele
NG_007529.1	FMR1 (2332)	Fwd	21812	G

Gene Model(s)

Function	mRNA				Protein		
	SNP to mRNA	Accession	Position	Allele change	Accession	Position	Residue change
ncRNA	Fwd	NR_033700.1	643	NA ⇒ NA	NA	NA	NA
ncRNA	Fwd	NR_033699.1	643	NA ⇒ NA	NA	NA	NA
cds-synon	Fwd	NM_002024.5	643	CGG ⇒ CGA	NP_002015.1	138	R [Arg] ⇒ R [Arg]
cds-synon	Fwd	NM_001185082.1	643	CGG ⇒ CGA	NP_001172011.1	138	R [Arg] ⇒ R [Arg]
cds-synon	Fwd	NM_001185081.1	643	CGG ⇒ CGA	NP_001172010.1	138	R [Arg] ⇒ R [Arg]
cds-synon	Fwd	NM_001185076.1	643	CGG ⇒ CGA	NP_001172005.1	138	R [Arg] ⇒ R [Arg]
cds-synon	Fwd	NM_001185075.1	643	CGG ⇒ CGA	NP_001172004.1	138	R [Arg] ⇒ R [Arg]

Figure 1 Reverse primer as reported by De Geyter et al.[1]: GCTCAGCTCCGTTTCGGTTTCACTTCCGGT. Note that the sequence is in the positive strand in the 5′–3′ direction.

though several studies have shown that a target can be amplified with few mismatches, a single base mismatch at the 3′ end as well as a few mismatches in the middle or toward the 5′ end generally reduce the PCR efficiency, although in different degrees. Mismatches in the 3′ end affect the target amplification greater than mismatches toward the 5′ end or in the middle.[3,4]

Second, the analysis performed by primer 3 software showed high Tm for the F′ primer, as well as high hairpin stability. Although it is possible to amplify templates with primers with high Tm using a two-step PCR, the methodology specifically states that they used the GC-Rich PCR system's instructions (Roche Applied Sciences, Rotkreuz, Switzerland), which follow a traditional PCR protocol. This, combined with the calculated high hairpin stability, and the variant in the 3′ end could have had a negative effect on the PCR efficiency.[3,4] Because the conventional PCR was the first step in its three-step protocol,[1] and because it was the one that defined which samples would undergo a triplet PCR, this reduced efficiency would be undesirable.

Locus-specific primers are, arguably, the most important variable in a successful PCR. With the advent of new technological tools to generate and validate primers,[4] it is important that investigators use these tools to maximize the efficiency of PCR, especially in the first steps of the methodology, because many samples might be incorrectly characterized as normal, possibly compromising the results of the study. Furthermore, an adequate report of the primer's sequence and orientation is crucial because it helps in the reproducibility of the results, the validation of the study, and the use of the methodology in subsequent studies.

ACKNOWLEDGMENTS

This work is part of the project "Determination of Fragile X Syndrome in Patients With Autism in Bucaramanga," which is funded by an investigation grant provided by the Universidad Industrial de Santander, Bucaramanga, Colombia.

DISCLOSURE

The authors declare no conflict of interest.

Julieth A. Sierra-Delgado, MD[1], Viviana L. Pérez, MSc[1] and Clara I. Vargas, MD, MSc[1]

[1]Grupo de Investigación en Genética Humana (GENEHUIS), Departamento de Ciencias Básicas, Facultad de Medicina, Universidad Industrial de Santander, Bucaramanga, Colombia. Correspondence: Julieth A. Sierra-Delgado (andreasierrad@outlook.com)

REFERENCES

1. De Geyter C, M'Rabet N, De Geyter J, et al. Similar prevalence of expanded CGG repeat lengths in the fragile X mental retardation I gene among infertile women and among women with proven fertility: a prospective study. *Genet Med* 2014;16:374–378.
2. Untergasser A, Cutcutache I, Koressaar T, et al. Primer3–new capabilities and interfaces. *Nucleic Acids Res* 2012;40:e115.
3. Sipos R, Székely AJ, Palatinszky M, Révész S, Márialigeti K, Nikolausz M. Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. *FEMS Microbiol Ecol* 2007;60:341–350.
4. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 2012;13:134.

Advance online publication 14 July 2016. doi:10.1038/gim.2016.78