Sections adapted from Alexandru Chelu's MSc thesis

Manual procedure for clinical-grade primer design

Primer pairs for exons within genes associated with Ehlers-Danlos syndrome were designed following the clinical protocol provided by the SDGS. The sequence of the gene of interest was downloaded from http://www.ensembl.org as whole genomic sequence, using the latest transcript available. Primer3Plus (Figure 7A) (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) was used to generate primer pairs inside the flanking intronic regions of each exon. Primer pairs were validated using SNPCheck V3 (https://genetools.org/SNPCheck/snpcheck.htm) to detect the presence of SNPs (Figure 7B). Clinical-grade primer pairs were selected if the amplicon size was below 500 bp and if SNPs were absent or filtered out. SNPs were replaced with an "N" in the original sequence and new primer pairs were generated and validated again.

Automatic clinical primer design using OmniPrimer

OmniPrimer was entirely developed in Python 3.8.2 to automatically perform each individual step of the clinical protocol described in the previous section. Selenium package 3.141.0 and built-in Python modules were used for the development of OmniPrimer. To run successfully, OmniPrimer has four prerequisites: (I) the operator should provide the file containing the sequence of the gene of interest in a text file. This must be placed in the main folder of OmniPrimer; (II) the operator should provide the name of the file; (III) the operator should provide the Uniform Resource Locator (URL) of the gene of interest from http://www.ensembl.org; (IV) a web browser should be installed on the device, which must be connected to the internet. The documentation of each primer pair is automatically downloaded as OmniPrimer is running and can be accessed by the operator immediately.

Manual primer design for individual exons

Fibrillin1 is an important extracellular matrix glycoprotein encoded by the *FBN1* gene, whose mutation has been associated with Ehlers-Danlos syndrome. During the digital work-placement in conjunction with the SDGS at the Sheffield's Children hospital, a total of 86 different clinical-grade primer pairs were generated to amplify and sequence all exons within *FBN1*. 21 out of 86 primer pairs covered the last exon of the gene, which was split in 21 parts due to its excessive length (3067 bp). Due to the abundance of SNPs within the genomic sequence, the majority of primer pairs were selected after numerous attempts (Figure 1). The most conserved regions of the gene could be easily processed within one or two attempts, whereas the variable regions were more time-consuming with multiple attempts required. Exons 16 and 53 were the most challenging exons to process, with 10 and 12 attempts respectively (Figure 1). All exons of *FBN1* were processed after over 17h of work, with an average of 10 minutes per primer pair designed. The work carried out on *FNB1* alone saved approximately 2 days of full-time work to clinicians, which translated into approximately £310 saved to the NHS. Moreover, about 4000 additional primer pairs were designed for 2071 different exons by 43 Human and Molecular Genetics MSc students, for a total of 1052 hours of work. Collectively, our joint efforts saved almost a month of work to clinicians and approximately £100,000 to the NHS.

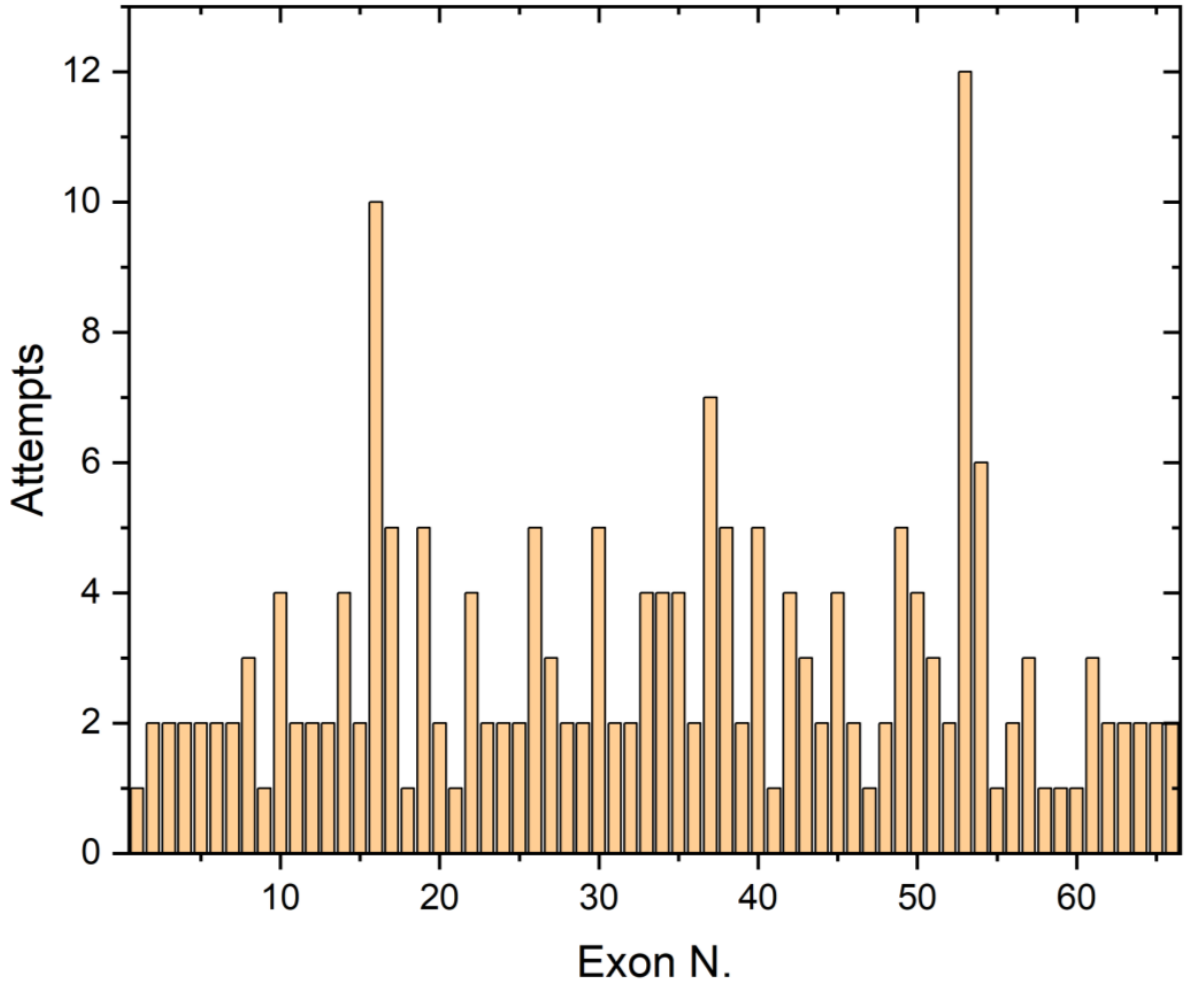


Figure 1. *FBN1* primer designing. The figure represents the attempts needed to design primer pairs for each exon in the *FBN1* gene. Primers in the most conserved regions were designed after 1 or 2 attempts whereas several attempts were necessary for variable regions. Data was plotted with Origin 2020b v.9.7.5.184.

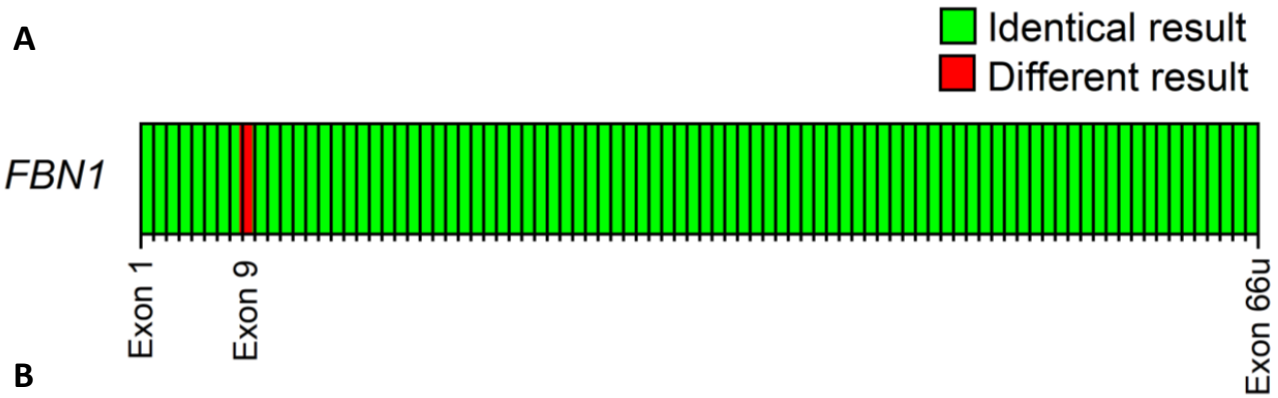
AUTOMATIC PRIMER DESIGN FOR COMPLETE GENES

Designing clinical-grade primer sets is currently done manually by clinicians and it requires an extensive time. The main factor that makes the protocol time-consuming is the numerous attempts to obtain primer sets within genomic regions that are highly repetitive and/or SNP-rich. This was evident during the work carried out on *FNB1* (Figure 1). Due to the digital nature of the protocol, a novel bioinformatic tool called OmniPrimer was developed with the purpose to automate each step of the protocol. OmniPrimer is capable to generate clinical-grade primer sets following a cascade of instructions. A multi-step algorithm was designed to process the genomic sequence within the file provided. The algorithm performs the following actions: (I) full extraction of exonic and intronic sequences; (II) splitting of exons bigger than 320 bp into the minimum number of parts equal in size; (III) 5' and 3' sequence flanking of each exon and exon part.

A library of fully flanked exons is generated at the end of the algorithm. Each flanked exonic sequence is then automatically copied in Primer3Plus by OmniPrimer, which prompts primer generation. Each newly generated primer pair is then copied and analyzed in SNPCheck for SNP detection. If a primer pair does not contain SNPs and the amplicon is smaller or equal to 500 bp, then the files associated with that primer pair are downloaded by OmniPrimer. Alternatively, if SNPs are found they are automatically replaced by "N"s and a new attempt is performed by the program. By looping into the exon library, a primer pair is generated for

every exon found, until no exons are left. OmniPrimer was initially tested on *FBN1* to prove that its results are identical to those of a human being. Each primer pair generated by OmniPrimer was compared to the corresponding one designed manually, hence using then as controls. 85 out of 86 primer pairs generated by OmniPrimer were identical to the controls, showing a total accuracy of 98.8% (Figure 2A). In the single occurrence where the results were different, OmniPrimer generated a primer pair with no SNPs, whereas the control contained 1 filtered out SNP (Figure 2B, C). Thus, OmniPrimer designed a primer pair qualitatively superior to a human being.

A



B

SNPCheck Results

Overview

Name	Primer 1	Matching/ Mismatching bases	Primer 2	Matching/ Mismatching bases	Chr.	Amplified Region	Result	SNPs
FBN1_Ex9_F-FBN1_Ex9_R	tgagtccttctactgacgaatgg	23/0	tggaactgacttacacaaacca	22/0	15	246 bp 48818279.. 48818524	All Filtered Out * 1 filtered	rs11384803 9

C

SNPCheck Results

Overview

Name	Primer 1	Matching/ Mismatching bases	Primer 2	Matching/ Mismatching bases	Chr.	Amplified Region	Result	SNPs
OMNI_FBN1_Ex9_F-FBN1_Ex9_R	gctgtttccagggacatgat	20/0	tgttctttaagagggcagtcaa	22/0	15	397 bp 48818166.. 48818562	No SNPs Found	

Figure 2. OmniPrimer validation on *FBN1*. OmniPrimer was tested on *FBN1* to prove that its results are reliable. **(A)** Comparison between primer pairs designed manually and automatically with OmniPrimer across 86 total cases. **(B)** Primer pairs designed manually for Exon 9 contained 1 filtered out SNP. **(C)** Primer pairs designed automatically for Exon 9 with OmniPrimer contained no SNPs. Data from figure A was plotted with Origin 2020b v.9.7.5.184.

To prove than OmniPrimer can process any gene and quantify the processing time, 10 different genes from 4 different inherited diseases were used as validation candidates (Figure 3). Designing primers for a complete gene using the program was achievable in minutes, with the lowest time being 8 minutes for *ENG* (23 primer pairs) and the highest being 2h and 52 minutes for *SMAD2* (229 primer pairs) (Figure 3).
Contrastingly, it was esteemed that several hours are required to accomplish the same task manually (Figure 3). Remarkably, OmniPrimer demonstrated to be from 9 to 23 times faster to

design primer pairs compared to a human being. The program was not capable to design primer pairs neither for exon 12 in *PIEZO2* nor for exon 1 and 3 parts of exon 11 in *SMAD2*. Therefore, unlike the other genes, the total coverage of *PIEZO2* and *SMAD2* was 98.5% and 98.3% respectively, which remains extremely high. Taken together, OmniPrimer proved to be an accurate, robust and time-saving solution to design primer in an automated manner.

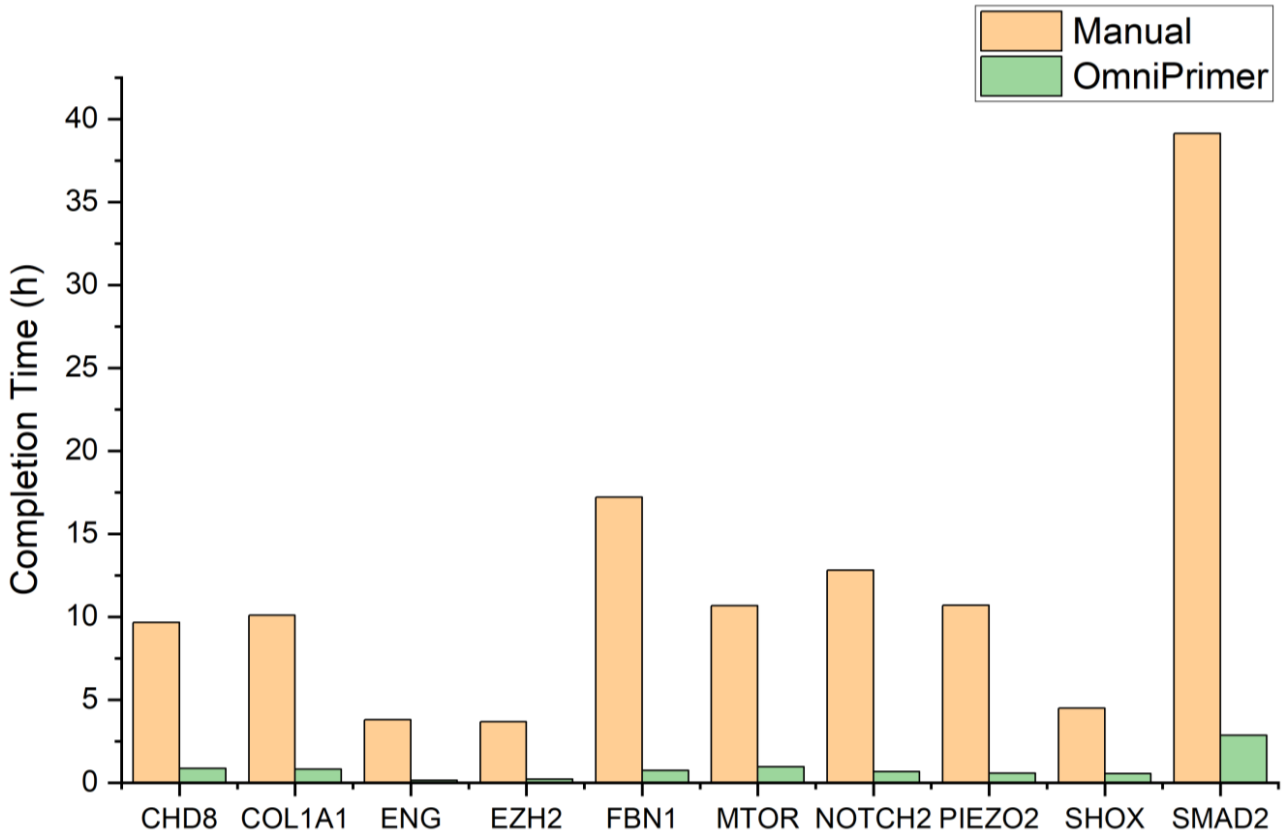


Figure 3. Quantification of time elapsed for complete gene coverage. Primer pairs were designed for 10 full genes by OmniPrimer. Full gene coverage can be achieved on average in 50 minutes by the program. The manual completion time was esteemed assuming that 10 minutes are required for every primer pair designed. Data was plotted with Origin 2020b v.9.7.5.184.