

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Primer Design to Sequence Analysis - a Pipeline for a Molecular Genetic Diagnostic Laboratory

Elaine Doherty¹, Renate Marquis-Nicholson¹, Jennifer M. Love¹,
Clare Brookes¹, Debra Prosser¹ and Donald R. Love^{1,2}

¹*Diagnostic Genetics, LabPlus, Auckland City Hospital, PO Box 110031, Auckland 1148,*

²*School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland 1142
New Zealand*

1. Introduction

Confirmation of a clinical diagnosis of a heritable disorder usually involves the analysis of the coding exons of a gene, plus surrounding intronic sequences. Variants that are identified using this strategy can be classed as either disease-causing or non-pathogenic polymorphisms. The variants comprise point mutations, micro insertions and deletions (indels), and copy number changes.

The detection of small variants (such as substitutions and indels) can be identified using a number of techniques including SSCP (Orita et al., 1989), DHPLC (Yu et al., 2006) and high-resolution melt-curve analysis (Garritano et al., 2009; Montgomery et al., 2010; Ngyyen-Dumont et al., 2009). These techniques can be difficult to optimise and can result in both false negative and false positive results. The accepted gold standard of mutation detection, however, is dideoxy-based sequencing. This is the most direct approach and can be the most sensitive of all techniques. However, it is also the most expensive option and can be the most time-consuming due to bottle-necks arising from designing appropriate primers, the analysis of subsequent sequence data, and the interrogation of data to filter out sequence variants that are not disease-causing. Within our laboratory, we experienced difficulties in three principal areas: standardising the primer design; objectively evaluating sequence data that can be performed in an automated manner thereby addressing the bottle-neck of trace sequence evaluation; and responding to the increasing need for in-depth analysis of sequence variants to inform referring physicians as to the clinical significance of the data.

To resolve the above issues, we developed a system that addressed each of these areas. The first involved designing primers that are assessed for unique amplification of targeted genomic regions using standard conditions. The aim here was not to tailor PCR conditions to a primer pair, but to force all primer designs to satisfy only one condition. The ancillary design feature of this module was to tail all primers such that all amplicons could be sequenced bi-directionally using only two primers (the “tail” sequences), thereby avoiding sequencing using amplicon-specific primers. The second area involved the wholesale adoption of Variant Reporter™ software (Applied Biosystems Ltd) for the automated analysis of all sequence data. This decision led to a significant initial investment in time to establish appropriate reference sequence projects for each of the genes analysed by our

laboratory. But critically, these templates can be applied to any gene of interest and have been successfully implemented within our laboratory to produce quality trace data for use in a diagnostic testing. The third area comprised a streamlined approach to bioinformatic analysis of sequence variants, which has replaced *ad hoc* and superficial analysis in order to support our reports to referring physicians. This aspect of our work remains a constant challenge and should be viewed as work-in-progress as more web-based bioinformatic tools become available.

2. Materials and methods

2.1 DNA purification

DNA was extracted from peripheral blood (taken in EDTA) using the Gentra Puregene DNA Extraction kit (Qiagen Ltd). If blood spots/Guthrie spots were used then DNA was extracted using the Qiagen Miniblood kit (Qiagen Ltd).

2.2 PCR conditions

PCR was performed using the following reagents in a 25µl reaction volume: 50 ng genomic DNA, Faststart buffer (Roche Applied Science), 2 mM MgCl₂, forward and reverse primers at 0.8µM each, 0.4mM dNTPs (GE Healthcare Ltd), 1 unit Fast Start Taq DNA polymerase (Roche Applied Science) and +/- 5µl GC-rich solution (Roche Applied Science); most PCRs did not require GC-rich solution.

The reaction conditions for all PCRs were: 95°C for 4 minutes followed by 35 cycles of 94°C for 45 seconds, 60°C for 30seconds and 72°C for 30seconds, with a final extension of 72°C for 10 minutes.

Primers designed using the Primer3 programme of the UCSC genome browser were tested first under the conditions above. If PCR failed, a second set of primers was designed using Primer3 and manual adjustment of the region to be amplified, or the alternative software design tool, FastPCR (see later).

5 µl of PCR product was purified using ExoSAP-IT® (USB) and subjected to bidirectional sequence analysis using Big-Dye Terminator v3.0 (Applied Biosystems), and 2.5ng per 100bp of amplified product. 20µl sequenced products were purified using Clean-Seq (Agencourt). 15ul of purified sequenced product was then subjected to capillary electrophoresis using a 3130x1s Applied Biosystems Genetic Analyser with a 50cm capillary array.

2.3 Sequence analysis

All trace sequences were analysed with KB™ Basecaller v1.4. Most of the sequence analysis parameters were set using the default values defined in Variant Reporter™ (Applied Biosystems). The mixed base threshold was set at the default of 25%. After base calling, the 5' and 3' ends were trimmed to remove sequence until fewer than 4 bases in a window of 20 had Quality values of less than 20; the resulting sequence is termed the clear range or contiguous read length (CRL). Traces were filtered with a minimum Trimmed Read Length (TRL) of fewer than 50 base pairs; the TRL is sequence that remains after trimming using the defined quality values. Traces were filtered with a Median Under Peak (PUP) of less than 4 (this is the median value within the clear range of the ratio of the signal of the highest secondary peak to the signal of the main called base). Traces were filtered if the expected TRL was less than 50% of the trace length providing a minimum trace score of 35. This is the average quality value of bases in the clear range which equates to a base-call error rate of

0.031%. A reference negative control DNA sample was sequenced for each gene in order for the algorithm to work.

3. Results

3.1 Primer design

Our main aim was to produce high-quality sequence traces spanning all coding exons plus flanking intronic sequence of any gene of interest. We arbitrarily chose to examine 20 base pairs of flanking intronic sequence in order to detect splice acceptor and donor variants. The priority for the primer design process was to produce primers that allowed for PCR amplification under standard buffer conditions using only one cycling protocol. Placing the emphasis on primer design rather than PCR optimisation was advantageous in reducing both technical time and reagent costs. Figure 1 shows the primer design workflow.

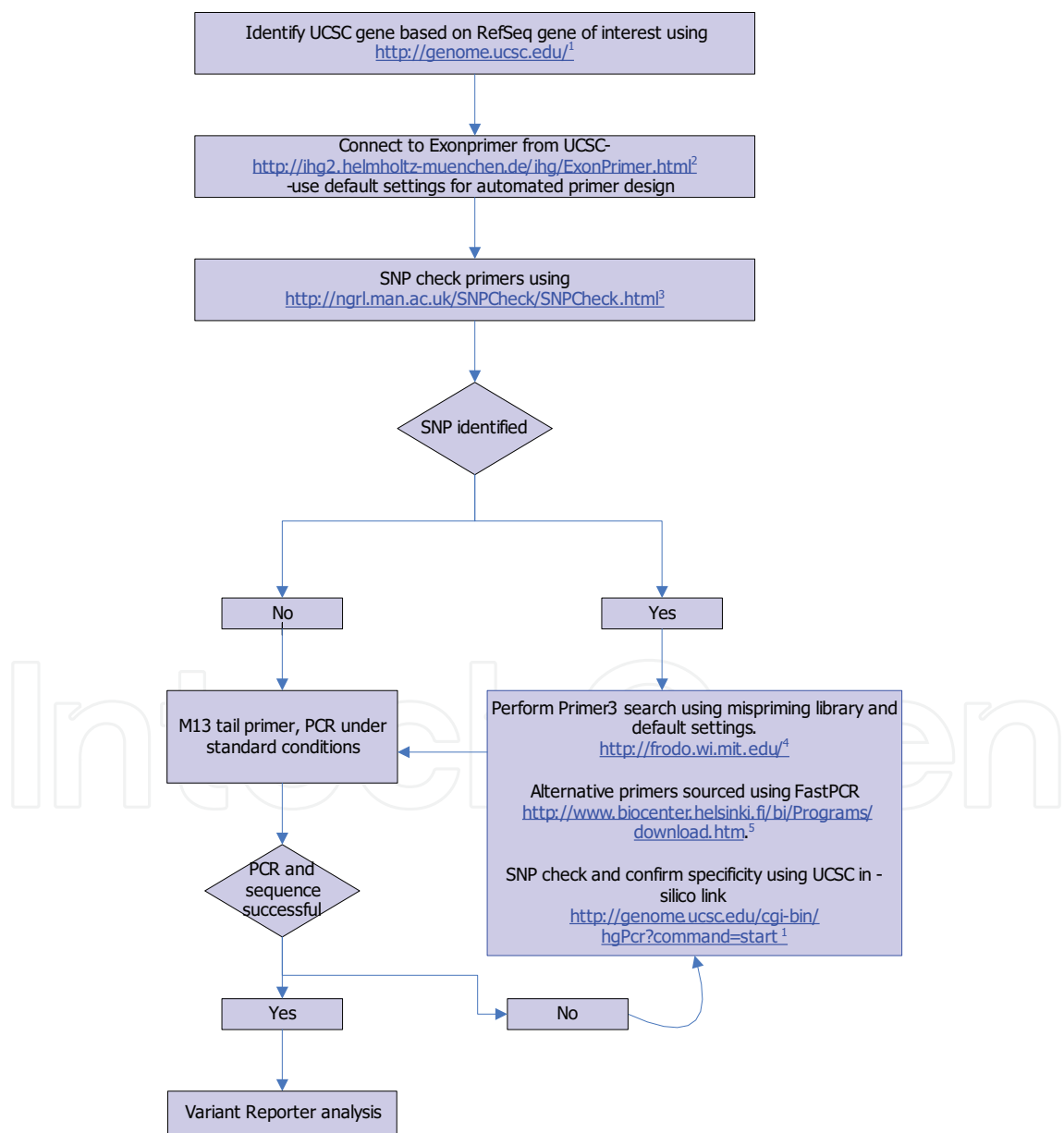


Fig. 1. Primer design workflow

The first objective involves the identification of the mRNA sequence of each gene of interest. This is achieved by interrogating the public UCSC genome browser page <http://genome.ucsc.edu/> (Figure 2A) and identifying an appropriate RefSeq accession number (Figure 2B), and the relevant UCSC accession number (Figure 2C). This website also provides a direct link to ExonPrimer for design of primers specific to this mRNA sequence (<http://ihg.helmholtz-muenchen.de/ihg/ExonPrimer.html>); see Figure 2. ExonPrimer provides exon and amplicon size and creates additional primer sets as alternatives. We use the ExonPrimer default parameters as this achieves our objective of designing primers flanking the 20 bases of intronic sequence.

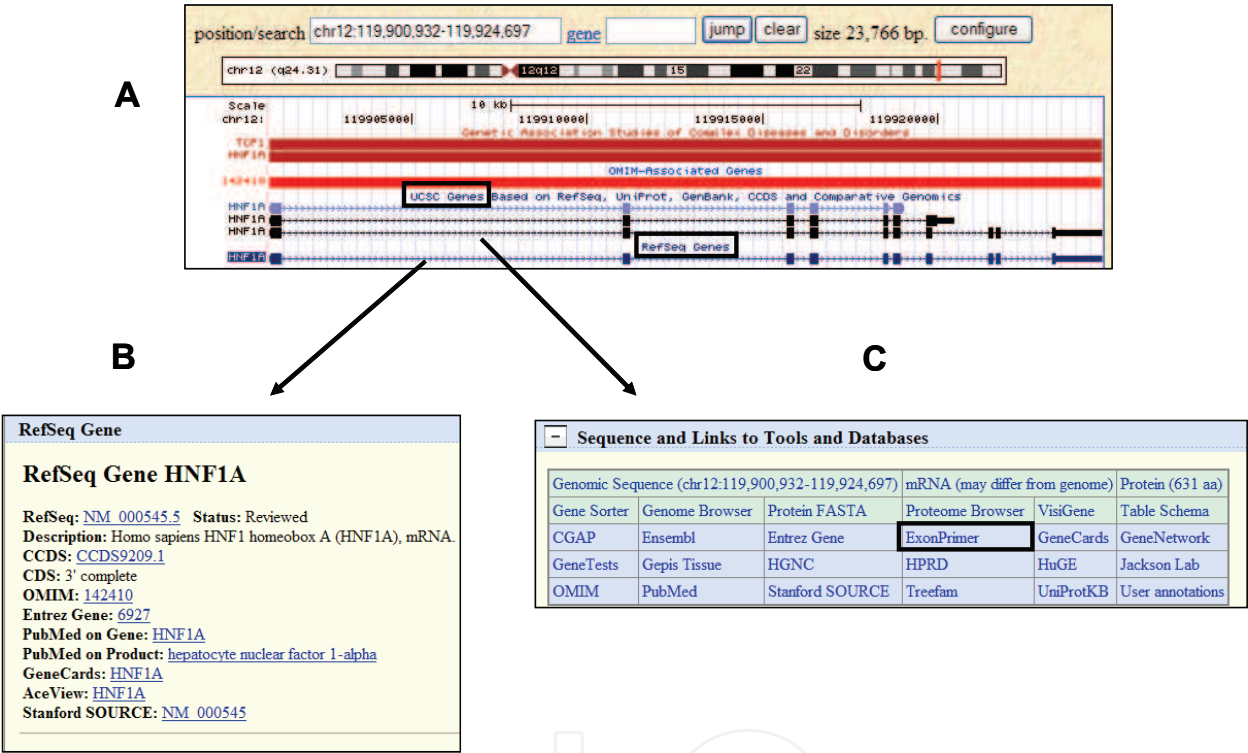


Fig. 2. ExonPrimer design

Following primer design, all primers are checked for single nucleotide polymorphisms (SNPs) using the software tool available from the National Genetic Reference Laboratory, Manchester, UK (<https://ngrl.manchester.ac.uk/SNPCheckV2/snpcheck.htm>). This is a bioinformatics program which allows for batch checking of primers for SNPs. It uses the current build of the human genome to identify the position in the sequence where the primers bind. The current release of dbSNP is used to detect any known SNPs at the primer binding sites. Primers that do not show SNPs are tailed with universal M13 sequences and used in PCR +/- GC-rich solution and our standard amplification conditions. Figure 3 shows a schematic of the tailed primers flanking two exons of a hypothetical gene (panel A), and the amplification products achieved after the PCR amplification of 11 unique exons (+/- genomic DNA).



Fig. 3. Schematic of amplicons and gel electrophoresis of amplicons

If a SNP is detected then a second set of primers is designed using Primer3 (Rozen & Skaletsky et al., 2000; <http://frodo.wi.mit.edu/primer3/>) in which we use the default settings; however, we highlight the human mis-priming library to avoid designing primers to regions of interspersed repeats. If the need arises, alternative primers are also designed using the freely available FastPCR programme (<http://www.biocenter.helsinki.fi/bi/Programs/download.html>). This programme includes various bioinformatics tools that can be used for the design of a variety of PCR applications. Following identification of primers using the default settings, we select primer pairs with annealing temperatures close to 60°C and amplicon lengths of approximately 400bp for optimal electrophoresis through our 50cm capillary array. Again, SNP checks of these primers are performed and all successful primers are tailed with universal M13 forward or reverse sequences. Unlike Exonprimer, these tools do not have an integrated search for cross-complementation to other areas of the genome. The UCSC website can be used for this purpose by selecting the *in silico* PCR option. This screens the primer pairs against the human genome and provides details of sequences that lie between the primers, and includes the primer pair. It also includes chromosome location and hence can be used to identify primer pairs which amplify homologous regions of the genome. The above design-on-demand pipeline has enabled us to design and implement primers for screening a number of disease-associated genes; we are in the process of writing a programme that will perform all of the above steps automatically.

3.2 Variant Reporter™ Reference sequence design

Our chosen software for the analysis of sequence traces is Variant Reporter™ developed by Applied Biosystems. This software tool uses advanced algorithms and quality metrics to automate the detection of variants and streamline the data analysis process, thereby greatly reducing the time required to analyse trace sequences. Additional features also allow for quality trace scoring over defined regions.

The first step in setting up sequence analysis using Variant Reporter™ is to produce a reference sequence for each gene, as shown in figure 4.

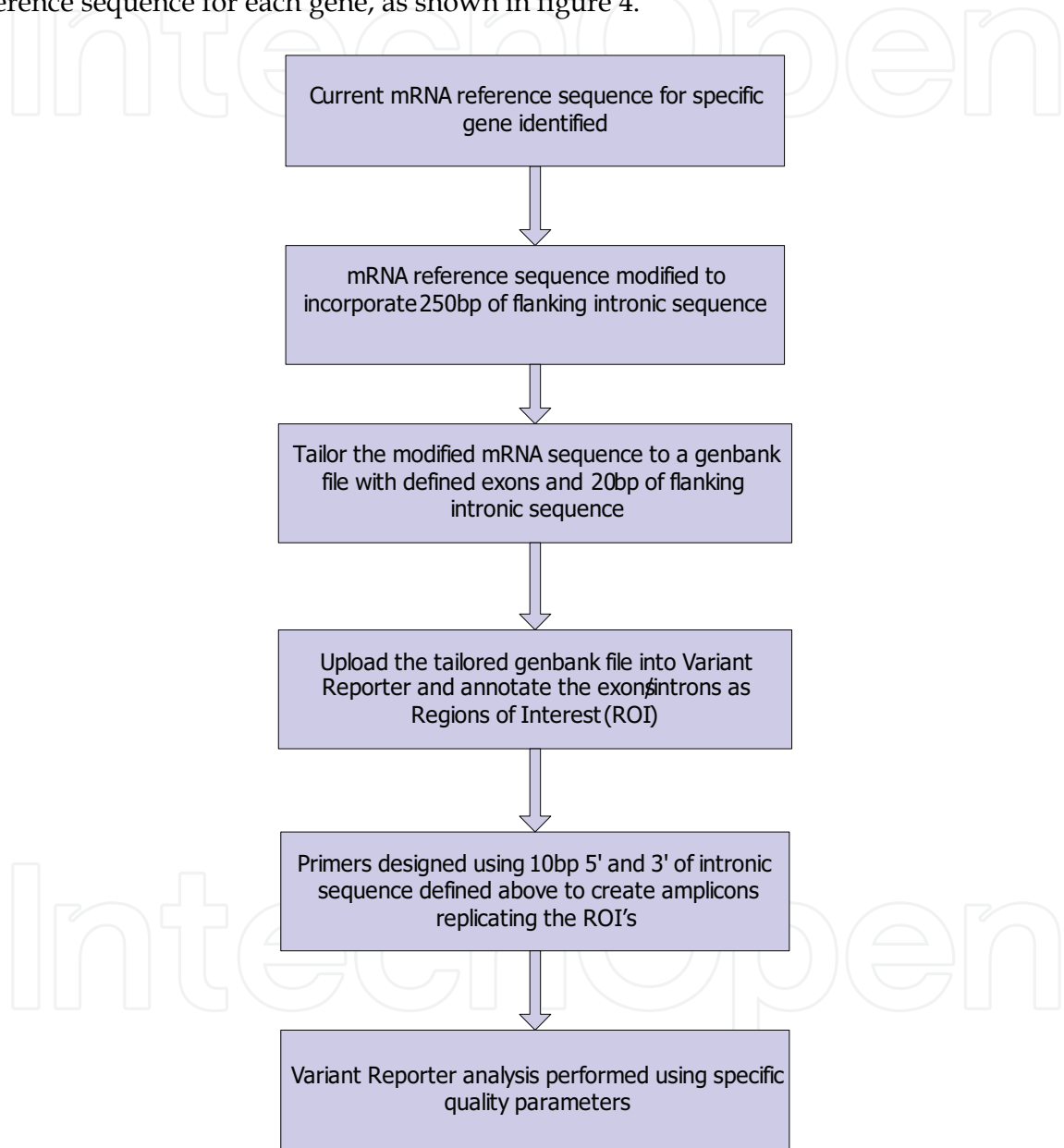


Fig. 4. General workflow in setting up a project in Variant Reporter™

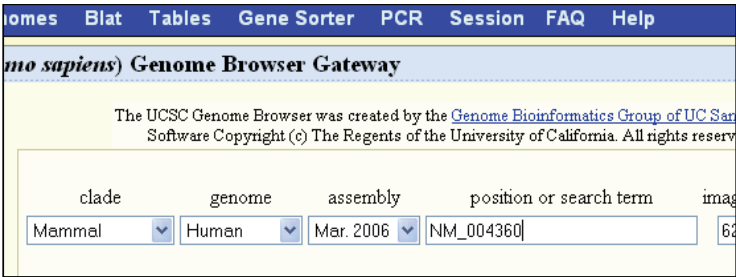
In our experience, the Variant Reporter™ programme has difficulty accepting large genomic reference sequences, therefore we reduce the reference sequence to a size which can be accommodated by the software. In our laboratory the reference sequences used are those of the mRNA of the gene of interest, which can be downloaded from the UCSC genome browser. This approach is in accordance with current international guidelines, which specify the reporting of

changes in relation to the mRNA accession number as specified by the Human Genome Variation Society (HGVS) Guidelines (<http://www.genomic.unimelb.edu.au/mdi/mutnomen/>). We modify each mRNA reference sequence using the UCSC genome browser to include enough intronic sequence to cover the maximum length of every amplicon we intend to sequence. This involves amending the mRNA reference sequence of the gene of interest to include 250bp of 5' and 3'intronic sequence. We then adapt the modified mRNA reference file to a tailored Genbank reference file which includes the defined exons and 20bp of flanking introns to create a "pseudo Genbank" file. Once the file has been imported into Variant Reporter™ it can be annotated to define the regions of interest (ROIs). In the case of small genes, the workflow for constructing appropriate files for Variant Reporter™ is described below.

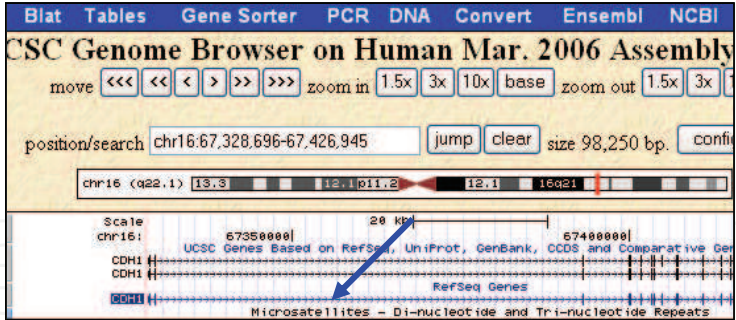
1. Identify the mRNA reference sequence for the targeted gene

The first stage is to identify the current mRNA reference sequence for the gene of interest in the UCSC genome browser. This is done by entering the UCSC genome browser and clicking on the 'Genomes' link and entering the gene name or RefSeq accession number in the search term box (Figure 5A). On the next page, click on the highlighted RefSeq (Figure 5B), and scroll to the 'Links to sequence' and click on 'Genomic sequence' (Figure 5C).

A



B



C

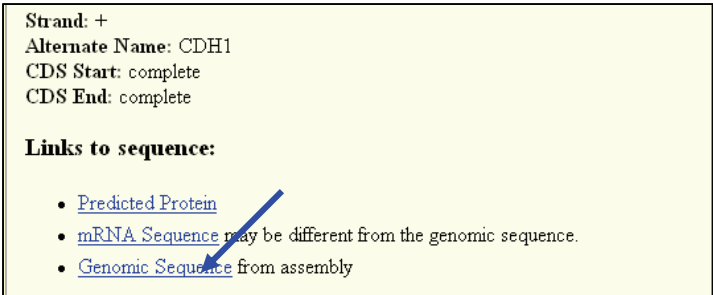


Fig. 5. Retrieving genomic sequence of gene of interest

2. Modify the mRNA reference sequence to incorporate 250bp of flanking intronic sequence.

The Sequence Retrieval Region Option is used to set the regions to 250bp (or to the user specification; Figure 6A). The Sequence Formatting Options allows for exon formatting in upper case and intronic regions in lower case (Figure 6B). This formatting is extremely useful for downstream visualisation of these regions in Variant Reporter™. The resulting sequence can be pasted into an appropriate program (such as notepad) in which all the RefSeq annotations are removed and the file is saved as .txt file. The lengths of each exon can be calculated easily using the word count character option. At this stage the modified mRNA sequence can be uploaded into Variant Reporter™. The exons and flanking intronic sequence can then be manually annotated.

A

HomeGenomesGenome BrowserBlatTablesGene SorterPCRSessionFAQHelp

Genomic Sequence Near Gene

Get Genomic Sequence Near Gene

Note: if you would prefer to get DNA for more than one feature of this track at a time, try the [Table Browser](#) using the output format s

Sequence Retrieval Region Options:

☐ Promoter/Upstream by bases

☐ 5' UTR Exons

☒ CDS Exons

☐ 3' UTR Exons

☐ Introns

☒ Downstream by bases

☐ One FASTA record per gene.

☒ One FASTA record per region (exon, intron, etc.) with extra bases upstream (5') and extra downstream (3')

☐ Split UTR and CDS parts of an exon into separate FASTA records

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated the edge of the chromosome.

Sequence Formatting Options:

☒ Exons in upper case, everything else in lower case.

☐ CDS in upper case, UTR in lower case.

☐ All upper case.

☐ All lower case.

☐ Mask repeats: ☒ to lower case ☐ to N

B

ECadherin_UCSC_250pad.txt - Notepad

FileEditFormatViewHelp

```
gctccgccctggggaggggtccgcgctgctgattggctgtggccggcagg
tgaaccctcagccaatcagcgggtacggggggcggtgcctccggggctcac
ctggctgcagccacgcacccctctcagtgccgtcggaactgcaaagcac
ctgtgagcttgccgaagtcagttcagactccagccgctccagccggcc
cgaccgcacccgcgcctgcctcgctcgccgtccccggccagcc
ATGGGCCCTTGGAGCCGAGCCTCTCGGCGCTGCTGCTGCTGCAAGgt
accccgatcccctgacttgaggggacgcattcgggccgcaagctccgc
gccccagccctgcgcccctcctctcccgctcgtcacccgttcccttcttc
caagaaagttcgggtcctgaggagcggagcggcctggaagcctcgcgcg
tcgggacccccagtgatgggagtgggggtgggtggtgaggggcgagcg
cggtttcctgccccctccagcgcagaccgagggcggtctggcc
>hg18_refGene_NM_004360_1 range=chr16:67329451-67330065 5'
aagggaacggtgggctaggtcttgagggggtgactcggcgccccctccc
gggagttcctgggggctcggcgccgtaggtttcgggggtgggggaggggtg
acgtcgctgcgcccgctcccggggctgcgggctgggggtcctcccccaat
cccgacgcgggagcgagggaggggcggcgctgttggtttcggtgagcag
gaggggaacctccgagtcacccggttccatctacctttccccacccccag
GTCTCTCTTGGCTCTGCCAGGAGCCGGAGCCCTGCCACCCTGGCTTTGA
CGCCGAGAGCTACACGTTACGGTGCCCCGGCGCCACCTGGAGAGAGGCC
GCGTCCTGGGCAGAGgtgagggcgcgctgccggtgtccctgggcggagtga
gggaggggttggaaggggccgagaaattgcactccacacccctgggtt
gcaatgggcaagctccctccttgggtcaaacgcaccccttgaatttac
gcagatttggggatccaaacgtttacgactgaacactgtggtggaggggg
taacctgcctggttgttgactatgtataaagaagaggaggttgagggc
acttactagaagcct
>hg18_refGene_NM_004360_2 range=chr16:67392824-67393547 5'
agttacttttagacaagataatattttcaattgacaagaagatgagaaa
gaaatcagagcacaagggaagtcacctgtagttgtgaagaacattatggt
gatagctttgttgtctgtcaccagattatgggttgggttgggtgtgttt
```

Fig. 6. Retrieving exon sequences with flanking 250bp of intron sequence

3. Annotate exons and introns as regions of interest (ROI)

The ‘define reference’ link of Variant Reporter™ is used to import the ‘modified reference sequence’ from step 2. At this point the user can choose to delete redundant reference sequences using the ‘delete layer’ option from the ‘Layers’ tab. The Variant Reporter™ reference sequence can now be annotated according to the user specifications using the ‘Create/Edit ROI’ option from the ROI tab. The ROI start for each exon is specified on the ‘Cumulative Exon start co-ordinates section’ of an exon formatting spreadsheet.

4. Design primers using 10bp 5' and 3' of intronic sequence to create amplicons replicating the ROI.

A main feature of Variant Reporter™ is that it provides quality statistics based on the amplicon, not the region of interest (ROI). Therefore, we adopted an *in silico* approach to make the ROIs coincident with the amplicons. This approach involved using the first 10 bases of each 5' intronic region as the forward primer, and reverse complemented the last 10 bases of the corresponding 3' intron to create the reverse primer. A freely available reverse-complement web-site (http://www.bioinformatics.org/SMS/rev_comp.html) can be used for this purpose. The primers can be easily identified by highlighting the corresponding bases in Variant Reporter™ and pasting them into a .txt file (Figure 7A).

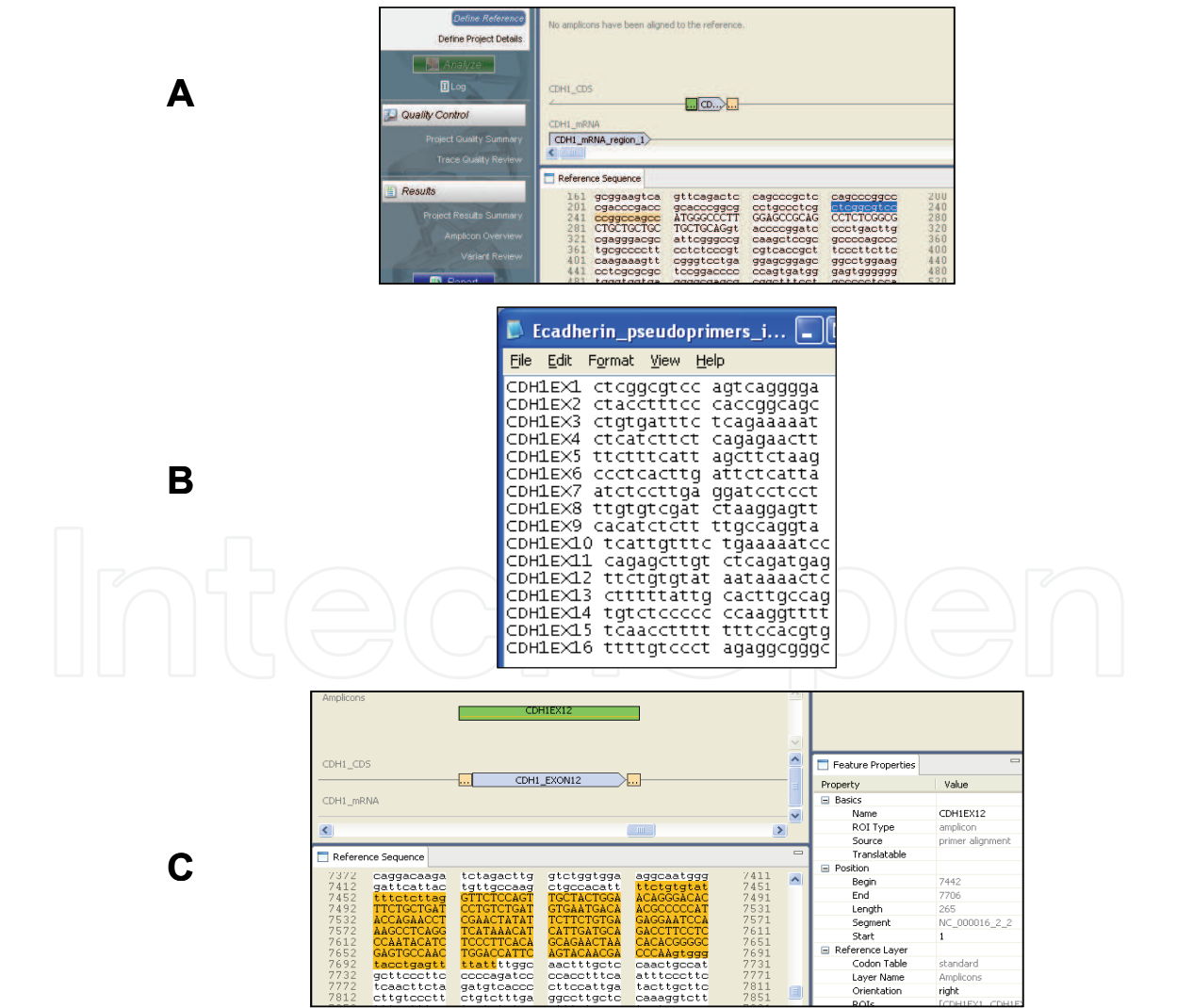


Fig. 7. Creation of regions of interest to replicate amplicons

The reverse-complement web browser is then used to reverse complement the 3' intron. An example of the .txt format that can be used to upload the primers into Variant Reporter™ is shown in Figure 7B. Note that the primer pair name will correspond to the amplicon name as it appears in Variant Reporter™. Once the primer sequences have been imported, the amplicons are automatically created and correspond to the exon plus flanking 20bp intron (Figure 7C). Nucleotide numbering of the coding regions starts with nucleotide 1 as A of the major initiation codon. Intronic sequences are referred to as -n or +n where -1 denotes the invariant G of the AG splice acceptor site, and +1 indicates the invariant G of the GT splice donor site, in accordance with HGVS nomenclature.

5. Establishing specific quality parameters in Variant Reporter™

The user can select default Analysis Parameters or customize their personal settings. We choose to set a minimum trace score of 35 equating to an average base-call error rate of 0.031%. These can be set using the Analysis Parameters tab.

The above workflow has been used to establish sequencing projects for a large number of genes, some of which are shown in Table 1.

Disease	Gene	RefSeq mRNA accession no.	Coding exons
Adrenoleukodystrophy	<i>ABCD1</i>	NM_000033.2	1-6
Gastric Cancer (E-Cadherin)	<i>CDH1</i>	NM_004360.2	1-16
Maturity Onset of Diabetes in the Young (MODY) type 2	<i>GCK</i>	NM_000162.3	1-10
MODY3	<i>HNF-1a</i>	NM_000545.4	1-10
Non-insulin dependent diabetes mellitus (NIDM) type I	<i>HNF-4a</i>	NM_000457.3	1-10
Non ketotic hyperglycinaemia	<i>GLDC</i>	NM_000170.2	1-25
Holocarboxylase synthetase deficiency	<i>HLCS</i>	NM_000411.4	4-12
Rett Syndrome	<i>MECP2</i>	NM_004992.3	1-4
Paragangliomas and Pheochromocytoma	<i>SDHB</i>	NM_003000.2	1-9
	<i>SDHC</i>	NM_003001.3	1-6
	<i>SDHD</i>	NM_003002.1	1-4
Myoclonus dystonia	<i>SGCE</i>	NM_001099400.1	1-11
Von Hippel-Lindau syndrome	<i>VHL</i>	NM_000551.2	1-3
Long QT Syndrome (LQT) type 1	<i>KCNQ1</i>	NM_000218.2	1-16
LQT2	<i>HERG</i>	NM_000238.2	1-15
LQT3	<i>SCN5A</i>	NM_198056.2	2-28

LQT5	KCNE1	NM_000219.3	3
LQT6	KCNE2	NM_172201.1	2
LQT7	KCNJ2	NM_000891.2	2
Myoclonus dystonia	SGCE	NM_003919.2	1-10
Familial adenomatous polyposis	APC	NM_000038.4	2-16
Multiple Endocrine Neoplasia (MEN) type 2A/2B	RET	NM_020630.4	10-16

Table 1. Genes for which Variant Reporter™ projects have been established

3.3 Bioinformatic analysis

Comparison of the sequence traces from a patient with those of a negative control individual in Variant Reporter™ results in the identification of sequence variants that differ from the relevant RefSeq sequences. The final step for a diagnostic laboratory is to use a method of “filtering” these variants to identify those that are benign (likely to be, but not necessarily, silent mutations at the amino acid level), and so provide a much more limited set of variants that may or may not be disease-causing. Our routine work flow is shown in Figure 8, which accesses a number of web-based analysis tools to interrogate the data in a formal manner. The reader can refer to many recent publications that highlight the work of others in classifying sequence variants in the context of their disease association (Abkevich et al., 2004; Sullivan et al., 2006; Lee et al., 2008; Arnold et al., 2009).

The work flow can be divided into a series of modules, which are summarised below:

1. Convert the locations of sequence variants in HGVS nomenclature (<http://www.hgvs.org/mutnomen/>) into predicted consequences on translation using the Mutalyzer web site (<http://www.mutalyzer.nl/2.0/>).
2. Interrogate sequence variants against SNP databases to determine if they represent known polymorphisms with associated rs identifiers. These variants are present in the general population and so would likely be non-pathogenic.
3. Interrogate sequence variants against disease-associated databases such as the Leiden Open Variation Database (LOVD; <http://www.lovd.nl/2.0/>) and the HGVS locus specific mutation databases (<http://www.hgvs.org/dblist/glsdb.html>).
4. Predict the functional consequences of predicted amino acid changes on protein function using evolutionary sequence conservation and physico-chemical considerations provided by web-based tools such as SIFT (http://sift.jcvi.org/www/SIFT_seq_submit2.html) and Polyphen-2 (<http://genetics.bwh.harvard.edu/pph2/>); Ng & Henikoff 2003; Tavtigian et al., 2008; Adzhubei et al., 2010; Flanagan et al., 2010.
5. Predict the effect of sequence variants on splicing (Walker et al., 2010).

The outcome of the above workflow provides laboratory staff with information that either strongly supports the conclusion that a sequence variant is a (benign) polymorphism, a disease-causing mutation, or of unknown clinical significance. The latter creates dilemmas for the laboratory, but usually leads to continuing family studies to determine if the variant co-segregates with the disease, or *in vitro*/modelling studies that necessitate research-based laboratory involvement.

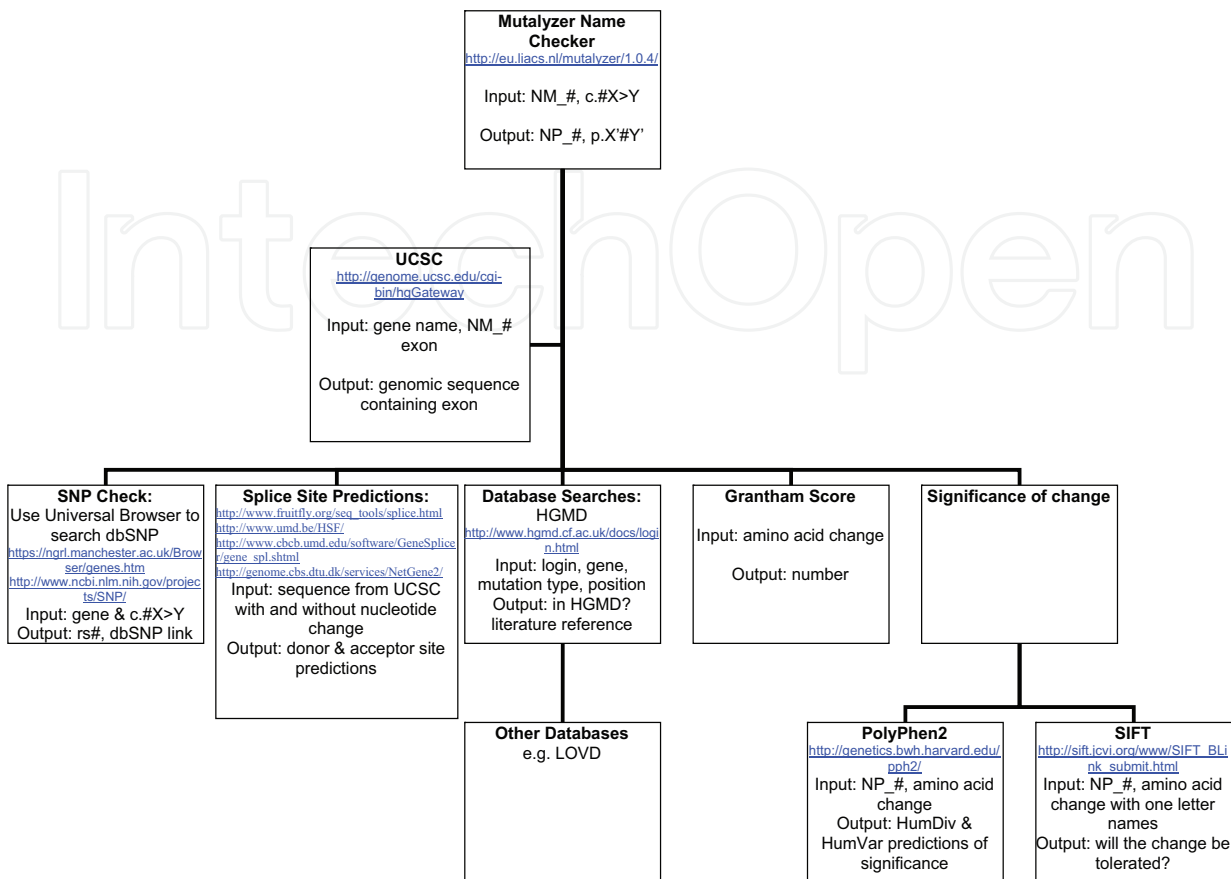


Fig. 8. Bioinformatic workflow

4. Discussion

The integration of the processes described above has allowed us to design a sequencing strategy suited to obtaining high-throughput quality controlled sequence data. To date, we have evaluated over 1000 sequence traces with the end point being high-quality data passing defined quality thresholds set in Variant Reporter™.

The first stage of the sequencing pipeline involved primer design. Here, our main emphasis was to create a structured template requiring minimal intervention from the operator. A tiered strategy consisting of the UCSC browser, Primer3 and FastPCR tools enables the selection of several primer pairs which can be M13-tailed and tested under standard conditions. By focusing initially on primer design rather than reaction optimisation we have reduced costs in technical time as well as reagents. In addition, M13 tailing of final-stage primers also decreases both operator time and error and opens the process to automation. In combination with the use of the SNP software tool and UCSC *in silico* PCR, this has enabled the rapid design of specific primers.

We chose to use primers which covered at least 20 bases of flanking introns in order to detect splice variants. Other laboratories may choose to design primers at different sites to

include deeper intronic variants. However, there is a lack of international best practice guidelines defining the length of intronic sequence necessary for diagnostic reporting. Until this is resolved, each laboratory will continue to sequence differing flanking lengths affecting the detection of deeper intronic variants that potentially have pathogenic properties. As recommended by the Clinical Molecular Genetics Society (CMGS; http://cmgsweb.shared.hosting.zen.co.uk/BPGs/Best_Practice_Guidelines.htm, retired publication), it is good practice for the diagnostic laboratory to specify the length of flanking intron that is sequenced in order for the clinician to gain a clear understanding of the genomic regions that have been assessed for each patient.

The second stage of the sequencing pipeline involved establishing Variant Reporter™ reference sequences and downstream analysis. This was initially run in parallel to our established sequencing analysis system, a more manual-orientated software tool which compared test and control sequences. Here, it was necessary for the user to scan the sequences as not all changes were automatically detected by the software. These variants were not genotyped by the software, therefore the user had to manually call the variant. In addition, the software did not provide quality scoring of the sequence data.

Initial problems were encountered as Variant Reporter™ was unable to accept large genomic reference sequences. Therefore we chose to create a modified reference sequence to a size that could be accommodated by the software. This approach involved using the UCSC genome browser to incorporate 250 bp of flanking intronic sequence into the selected mRNA sequence. In turn, this modified sequence could be annotated as a Genbank file and uploaded into Variant Reporter™ to create a reference file that included highlighted regions of interest. This process alleviated the time-consuming step of creating the reference sequence manually in Variant Reporter™.

At the point of analysis it is necessary to incorporate a reference sample in the project in addition to the test sample. We routinely use a pre-sequenced negative control to compare with patient samples. Without the use of this reference sample we have found that Variant Reporter™ fails to detect pre-identified variants. As described by Applied Biosystems, this is due to the algorithm requiring a second sample for trace comparison.

Once optimisation of the reference file and project parameters was completed, we analysed over 1000 sequence traces in parallel using Variant Reporter™ and our conventional method. All substitutions identified by our conventional method were successfully detected and genotyped automatically by Variant Reporter™. In addition, the heterozygosity/homozygosity of each variant matched. All data passed our quality threshold set in Variant Reporter™ of a minimum trace score of 35 equating to an average base-call error rate of 0.031% (as now specified on all diagnostic reports produced from our laboratory). Our new system has reduced the time taken to analyse samples to roughly 10% of the original time (a reduction from 8 hours for a full multi-gene screen to 1 hour).

A major inconsistency which we have identified, however, is the failure of Variant Reporter™ to identify and correctly name indels. In order to circumvent this problem we manually evaluate all samples not reaching 100% trimmed read length (TRL) as identified through the trace quality review using the procedures above. The % TRL is the length of the sequence that remains after trimming, divided by the amplicon length. A % TRL falling beneath the 100% threshold as identified through the trace score feature can be used to

indicate a potential insertion or deletion. Our in-house protocol for these samples now involves scanning of these traces in the variant review window.

Once an insertion/deletion mutation has been detected visually we manually score the result. This approach also addresses the problem of inconsistencies in genotyping data in that Variant Reporter™ uses a different nomenclature system from the current HGVS international guidelines for diagnostic laboratories (http://cmgsweb.shared.hosting.zen.co.uk/BPGs/Best_Practice_Guidelines.htm).

Both of these issues are currently being assessed by Applied Biosystems.

In our experience the Variant Reporter™ default trace score of 25 allows for trailing C shoulders indicative of a sub-optimal array. By increasing this score to 35, the threshold is raised sufficiently to exclude low-quality sequence data. At present there is no universal consensus on minimal quality settings that should be used by the genetic community. Until there is, each laboratory will continue to use its own standards of quality scoring which will result in different rates of false negatives for each laboratory. We feel that international guidelines are needed to address this concern within the diagnostic laboratory.

While we were implementing enhanced primer design and the assessment of sequence traces using Variant Reporter™, our bioinformatic analysis strategy also required attention. The approach described here is evolutionary as more programmes become available on-line to predict the biological significance of sequence variants. In this respect, it should be noted that similar challenges in biological interpretation are occurring in the field of whole genome molecular karyotyping (Marquis-Nicholson et al., 2010). The construction of international databases comprising well-curated data from diagnostic laboratories (Cotton et al., 2008) should improve our understanding of variants and improve the genetic analysis of individuals and their families with heritable disorders.

5. References

- Abkevich, V.; Zharkikh, A.; Deffenbaugh, A.M.; Frank, D.; Chen, Y.; Shattuck, D.; Skolnick, M.H.; Gutin, A. & Tavtigian, S.V. (2004). Analysis of missense variation in human BRCA1 in the context of interspecific sequence variation. *J Med Genet*, 41(7), 492-507.
- Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P.; Kondrashov, A.S. & Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. *Nat Methods*, 7(4), 248-249.
- Arnold, S.; Buchanan, D.D.; Barker, M.; Jaskowski, L.; Walsh, M.D.; Birney, G.; Woods, M.O.; Hopper, J.L.; Jenkins, M.A.; Brown, M.A.; Tavtigian, S.V.; Goldgar, D.E.; Young, J.P. & Spurdle, A.B. (2009). Classifying MLH1 and MSH2 variants using bioinformatic prediction, splicing assays, segregation, and tumor characteristics. *Hum Mutat*, 30(5), 757-770.
- Cotton, R.G.; Auerbach, A.D.; Axton, M.; Barash, C.I.; Berkovic, S.F.; Brookes, A.J.; Burn, J.; Cutting, G.; den Dunnen, J.T.; Flicek, P.; Freimer, N.; Greenblatt, M.S.; Howard, H.J.; Katz, M.; Macrae, F.A.; Maglott, D.; Möslin, G.; Povey, S.; Ramesar, R.S.; Richards, C.S.; Seminara, D.; Smith, T.D.; Sobrido, M.J.; Solbakk, J.H.; Tanzi, R.E.; Tavtigian, S.V.; Taylor, G.R.; Utsunomiya, J. & Watson, M. (2008). GENETICS. The

- Human Variome Project. *Science*, 322(5903), 861-862. Orita, M.; Iwahana, H.; Kanazawa, H.; Hayashi, K. & Sekiya, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as SSCPs. *Proceedings of the National Academy of Sciences of the United States of America (USA)*, 86(8), 2766-2770.
- Flanagan, S.E.; Patch, A.M. & Ellard, S. (2010). Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet Test Mol Biomarkers*, 14(4), 533-537.
- Garritano S, Gemignani F, Voegelé C, Nguyen-Dumont T, Le Calvez-Kelm F, De Silva, D.; Lesueur, F.; Landi, S. & Tavtigian, S.V. (2009). Determining the effectiveness of High Resolution Melting analysis for SNP genotyping and mutation scanning at the TP53 locus. *BMC Genet*, 10, 5.
- Lee, E.; McKean-Cowdin, R.; Ma, H.; Chen, Z.; Van Den Berg, D.; Henderson, B.E.; Bernstein, L. & Ursin, G. (2008). Evaluation of unclassified variants in the breast cancer susceptibility genes BRCA1 and BRCA2 using five methods: results from a population-based study of young breast cancer patients. *Breast Cancer Res*, 10(1), R19.
- Marquis-Nicholson, R.; Aftimos, S.; Hayes, I.; George, A. & Love, D.R. (2010). Array comparative genomic hybridisation: a new tool in the diagnostic genetic armoury. *N Z Med J*, 123(1318), 50-61.
- Montgomery, J.L.; Sanford, L.N. & Wittwer, C.T. (2010). High-resolution DNA melting analysis in clinical research and diagnostics. *Expert Rev Mol Diagn*, 10(2), 219-240.
- Ng, P.C. & Henikoff, S. (2003). SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 31(13), 3812-3814.
- Nguyen-Dumont, T., Calvez-Kelm, F.L., Forey, N., McKay-Chopin, S., Garritano, S., Gioia-Patricola, L., De Silva, D., Weigel, R., Sangrajang, S., Lesueur, F.; Tavtigian, S.V.; Breast Cancer Family Registries (BCFR); Kathleen Cuninghame Foundation Consortium for Research into Familial Breast Cancer (kConFab). (2009). Description and validation of high-throughput simultaneous genotyping and mutation scanning by high-resolution melting curve analysis. *Hum Mutat*, 30(6), 884-890.
- Rozen, S. & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, Krawetz, S. & Misener, S. (Eds), 365-386, Humana Press, Totowa, NJ.
- Sullivan, L.S.; Bowne, S.J.; Birch, D.G.; Highbanks-Wheaton, D.; Heckenlively, J.R.; Lewis, R.A.; Garcia, C.A.; Ruiz, R.S.; Blanton, S.H.; Northrup, H.; Gire, A.I.; Seaman, R.; Duzkale, H.; Spellicy, C.J.; Zhu, J.; Shankar, S.P. & Daiger, S.P. (2006). Prevalence of disease-causing mutations in families with autosomal dominant retinitis pigmentosa: a screen of known genes in 200 families. *Invest Ophthalmol Vis Sci*, 47(7), 3052-3064.
- Tavtigian, S.V.; Greenblatt, M.S.; Lesueur, F.; Byrnes, G.B.; IARC Unclassified Genetic Variants Working Group. (2008). In silico analysis of missense substitutions using sequence-alignment based methods. *Hum Mutat*, 29(11), 1327-1336.
- Walker, L.C.; Whaley, P.J.; Couch, F.J.; Farrugia, D.J.; Healey, S.; Eccles, D.M.; Lin, F.; Butler, S.A.; Goff, S.A.; Thompson, B.A.; Lakhani, S.R.; Da Silva, L.M.; kConFab Investigators, Tavtigian, S.V., Goldgar, D.E.; Brown, M.A. & Spurdle AB. (2010). Detection of splicing aberrations caused by BRCA1 and BRCA2 sequence variants

encoding missense substitutions: implications for prediction of pathogenicity. *Hum Mutat*, 31(6), E1484-505.

Yu, B.; Sawyer, N. A.; Chiu, C.; Oefner, P. J. & Underhill, P. A. (2006). DNA mutation detection using denaturing high-performance liquid chromatography (DHPLC). *Current Protocols in Human Genetics*. 7.10.1-7.10.14.

IntechOpen

IntechOpen



Applications and Experiences of Quality Control

Edited by Prof. Ognyan Ivanov

ISBN 978-953-307-236-4

Hard cover, 704 pages

Publisher InTech

Published online 26, April, 2011

Published in print edition April, 2011

The rich palette of topics set out in this book provides a sufficiently broad overview of the developments in the field of quality control. By providing detailed information on various aspects of quality control, this book can serve as a basis for starting interdisciplinary cooperation, which has increasingly become an integral part of scientific and applied research.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Elaine Doherty, Renate Marquis-Nicholson, Jennifer M. Love, Clare Brookes, Debra Prosser and Donald R. Love (2011). Primer Design to Sequence Analysis - a Pipeline for a Molecular Genetic Diagnostic Laboratory, Applications and Experiences of Quality Control, Prof. Ognyan Ivanov (Ed.), ISBN: 978-953-307-236-4, InTech, Available from: <http://www.intechopen.com/books/applications-and-experiences-of-quality-control/primer-design-to-sequence-analysis-a-pipeline-for-a-molecular-genetic-diagnostic-laboratory>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen