Toward prevention of Hemoglobinopathies in Oman

Hassan, S.M.

CHAPTER 12

MOLECULAR DIAGNOSTICS OF THE *HBB* GENE IN AN OMANI COHORT USING BENCH-TOP DNA ION TORRENT PGM TECHNOLOGY

Hassan SM, Vossen RHAM, Chessa R, den Dunnen JT, Bakker E, Giordano PC and Harteveld CL

Blood Cells, Molecules and Diseases. 2014;53:133–137

ABSTRACT

Hemoglobinopathies, such as sickle cell disease (SCD) and beta thalassemia major (TM), are severe diseases and the most common autosomal recessive condition worldwide and in particular in Oman. Early screening and diagnosis of carriers is the key for primary prevention. Once a country-wide population screening program is mandated by law, a sequencing technology that can rapidly confirm or identify disease-causing mutations for a large number of patients in a short period of time will be necessary.

While Sanger sequencing is the standard protocol for molecular diagnosis, next generation sequencing starts to become available to reference laboratories. Using the Ion Torrent PGM sequencer, we have analysed a cohort of 297 unrelated Omani cases and reliably identified mutations in the beta globin (*HBB*) gene. Our model study has shown that ion torrent PGM can rapidly sequence such a small gene in a large number of samples using a barcoded uni-directional or bi-directional sequence methodology, reducing cost, workload and providing accurate diagnosis. Based on our results we believe that the Ion Torrent PGM sequencing platform, able to analyse hundreds of patients simultaneously for a single disease gene can be a valid molecular screening alternative to ABI sequencing in the diagnosis of Hemoglobinopathies and other genetic disorders in the near future.

INTRODUCTION

Severe Hemoglobinopathies (HBP), such as sickle cell disease (SCD) and thalassemia major (TM) are the most common recessive disorders in Oman and β-thalassemia and sickle cell disease carriers are widely present in the country. A national premarital screening program for the detection of carriers for primary prevention of β-thalassemia major and sickle cell disease is available but not mandatory by law. Following carrier detection at the haematological and biochemical level, a high-throughput screening approach is needed to facilitate a population targeted molecular analysis program for HBP (or any other common genetic disorder) for rapid confirmation. New DNA sequencing technologies and platforms are continuously being updated to accommodate the fast growth of science and research trying to improve molecular methods and techniques that can lead to fast and reliable molecular diagnosis. Next-generation sequencing technologies have been offering reliable approaches to rapid DNA genotyping. These newly evolved technologies have demonstrated advantages over Sanger sequencing by generating megabases to gigabases of data, allowing direct detection of sequence variants (1) in a very short period of time. If these new sequencing platforms are to be used routinely for diagnosis purposes, a number of factors such as sample scalability and cost should be addressed. Ion Torrent's Personal Genome Machine (PGM) (Guilford, CT, USA; now owned by Life Technologies, Carlsbad, CA, USA), is one of the options available (2) since late 2010 (3) and capable of generating 100 Mb of sequence data on a 314 chip, within several hours of machine run time (4). Sequence data are obtained by directly sensing the ions produced by template-directed DNA polymerase synthesis. The ion chip contains ion-sensitive, field-effect transistor-based sensors in 1.2 million wells on a 314 chip, which can measure independent sequencing reactions (3). DNA from different patients can be evaluated on the same sequencing chip using the barcoding methodology (5), allowing simultaneous sequencing for hundreds of patients.

To test the feasibility of using the Ion Torrent PGM for clinical variant analysis, we assessed its performance to detect variants in the beta globin gene (*HBB*) responsible for beta thalassemia and sickle cell disease. Most next generation sequencing studies involve investigating regions of interest of a large number of exons (up to 50,000 exons), from ten to thousands of genes on few patients (6). Instead, we have used the same platform to screen a large cohort of patients of 297 patients simultaneously for a relatively small single gene (*HBB*). Ion Torrent's PGM proved to be a valuable tool in identifying variants in high throughput analysis. It is suitable for large scale diagnostic screening for such a common disease gene, responsible for sickle cell disease and beta thalassemia in Oman.

METHODS

Sample collection and experimental design

We performed ion torrent PGM sequencing on 297 unrelated individuals. These samples were preselected based on the traditional hematological tests, Cell Blood Count (CBC) and High Performance Liquid Chromatography (HPLC) patterns. HPLC detects HbS carriers based on the appearance of an abnromal fraction and a beta-thalassemia carrier by an elevated

HbA_2 fraction. Patients suspected to be carriers of beta thalassemia (n=137), beta thalassemia major (n=15) and sickle cell disease (n=132) as well as normal controls (n=13) were all involved in the molecular study. A signed consent form of each patient has been provided by their designated clinician. DNA was isolated from peripheral blood using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA, USA). To cover the common mutations, including the

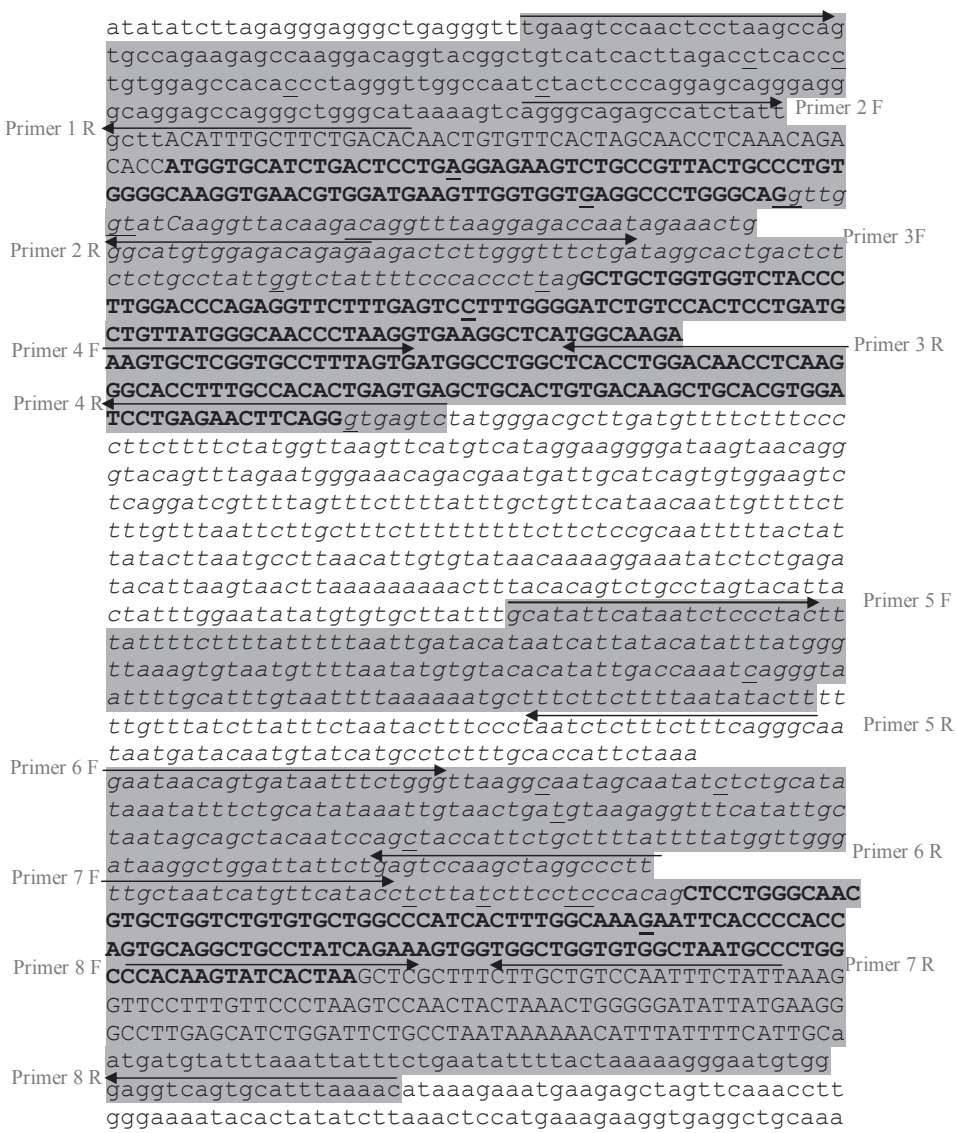

Figure 12.1. A schematic presentation of the beta globin gene (*HBB*). Exons are indicated in bold and introns in italics. The positions of the PCR primers are indicated as arrows above the sequence and the PCR products are highlighted. The frequently occurring mutations in the Omani population are underlined.

full coding regions, in the *HBB* gene, 8 separate PCR primer pairs were designed using the Primer3 oligonucleotide design tool (7). The primers sequences and the product size for each amplicon are listed in Table 12.1. The location of the primers within the *HBB* gene is depicted in Figure 12.1 covering the mutation spectrum present in the Omani population (11) that are underlined in the figure.

Ten different barcode primers consisting of 6 nucleotides were designed. For each fragment, the 5' end of the forward primer was tagged with M13 and the 10 barcodes giving 10 different primer sequences, whereas the 3' end of the reverse primer contained the P1-adapter sequence. Furthermore, the A-adapter - M13 primer was also tagged with the same 10 barcodes to enable indexing of a large number of patients using a relatively limited set of only 10 different barcodes (Figure 12.2); thus 10 barcoded target forward primers were designed to be combined with 10 barcoded A- adapter primers for a total of 100 (10x10) barcode combinations to perform

Table 12.1. Primer sequence of each fragment including the product size and the optimum annealing temperature.

Fragment	Primer	Fragment size	Tm
1	**Fw**: TGAAGTCCAACTCCTAAGCCA		
	Rv: GTGTCAGAAGCAAATGTAAGC	189 bp	55 °C
2	**Fw:** AGGGCAGAGCCATCTATT		
	Rv: TCTCTGTCTCCACATGCC	230 bp	55 °C
3	**Fw**: GAAGACTCTTGGGTTTCTGA		
	Rv: CTTGAGGTTGTCCAGGTGA	229 bp	56 °C
4	**Fw**: AAGTGCTCGGTGCCTTTAGTG		
	Rv: AGAAGGGGAAAGAAAACATCAAG	155 bp	55 °C
5	**Fw**: GCATATTCATAATCTCCCTAC		
	Rv: GCCCTGAAAGAAAGAGATTA	221 bp	55 °C
6	**Fw**: GAATAACAGTGATAATTTCTGGG		
	Rv: AAAGGGCCTAGCTTGGACTC	188 bp	55 °C
7	**Fw**: TGCTAATCATGTTCATACC		
	Rv: AATAGAAATTGGACAGCAAG	186 bp	53 °C
8	**Fw**: CCACAAGTATCACTAAGCTC		
	Rv: GTTTTAAATGCACTGACCTC	218 bp	53 °C

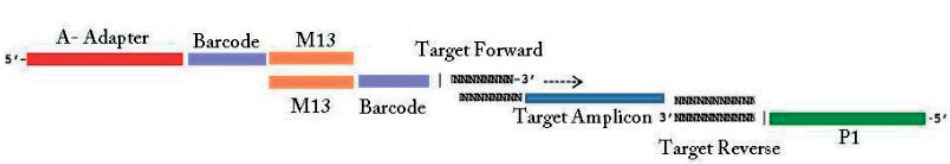

Figure 12.2. Schematic representation of the PCR showing the dual barcode design.

Table 12.2. List of the 10 barcode sequences used in this study model. The same 10 barcodes were tagged to the A-adapter-M13 primer and to the forward primer, giving (10x10) 100 different combination sequences for tagging/identification of 100 samples per run.

Barcode no.	Barcode sequence
1	GGTAAC
2	GAGAAC
3	GGATTC
4	AAGATC
5	AGGAAC
6	AAGTTC
7	TGATTC
8	GATAAC
9	CGGAAC
10	CCGAAC

analysis on 100 patients in a single run. The sequences of the barcodes are listed in Table 12.2. To ensure sequencing the total length of the amplified fragment, amplicon sizes were restricted to 230bp (including primers).

PCR conditions and library preparation

After PCR optimizations, amplifications were performed in a GeneAmp® PCR System 9700 (Applied Biosystems) in a total volume of 8µl PCR mix consisting of 1x PCR buffer, 1.8 mM $MgCl_2$, 200 µM dNTPs, 300 nM P1-adapter reverse primer, 300 nM A-adapter forward primer, 50 nM M13-target forward primer, 0.4 U FastStart High-Fidelity enzyme blend (5 U/µl, Roche Applied Science) and 10ng genomic DNA. Thermal cycling parameters were as follows: initial denaturation for 10 minutes at 95°C, followed by 35 cycles of 20s at 95°C, 30 s annealing at different temperatures (Table 12.1) and 40s at 72°C. Final extension was performed for 5 minutes at 72°C. The success rate of the amplifications was checked by adding 1µl of 20x EvaGreen® fluorescent dye (Biotium) to the wells of the PCR plate. By performing a melting analysis in the LC480 (Roche), the PCR yield of individual reactions could be visualised. PCR reactions deriving from 100 barcode-combinations were pooled according to their fluorescent levels after melting analysis. Size selection was done by isolating the correct band after electrophoresis on a 1.5% agarose gel. All fragments gave a single band except fragment 3 which had double bands. The correct band was cut out and the product was purified using the MinElute Gel Extraction Kit (Qiagen). The remaining pools were purified using Agencourt AMPure XP beads (Beckman Coulter) according to the manufacturer's instructions. The concentration and integrity of the library was determined by measuring the sample with a Bioanalyzer on a High Sensitivity DNA chip Agilent Technologies. The final library concentration that was used for emulsion PCR was set at 25pM.

Emulsion PCR, sequencing and data analysis

Emulsion PCR was performed using the Ion OneTouch™ 200 Template Kit v2 DL (Life Technologies) and the percentage of ion sphere enrichment was checked with a Qubit fluorometer (Life Technologies). Sequencing was done by using the Ion PGM Sequencing 200 Kit. Mutation analysis was done on patient-specific files using the NextGEN software version 2.3 (Softgenetics). Sanger sequencing was performed on few samples that were found to have a normal gene sequence in the carriers and one mutation in the homozygotes by ion torrent analysis.

RESULTS

Using our design, we were able to achieve overall very good coverage of the targeted regions. The coverage output for the first 100 patients is shown in Figure 12.3. Nearly identical coverage results were obtained from the other two runs. A subset of the base pair changes detected (Table12.3) were observed in all sequences of fragment 3 in controls and patients and therefore were excluded as artefacts. Seven patients that were suspected to carry one or two β-thalassemia mutation, revealed a normal sequence in the carriers and one mutation in the compound heterozygous diseased cases. These cases were re-sequenced by Sanger sequencing using a PCR with different primers and it confirmed the presence of one mutation in the *HBB* gene in the heterozygotes and two mutations in the compound heterozygotes. Four cases were found to carry the HbE mutation (Cd26 GAG>AAG), 2 in carriers and 2 in association with another β-mutation, and 3 cases to have the IVS-I-128 T>G mutation, 1 in a carrier and 2 in association with another β-mutation. To investigate why the Ion Torrent PGM sequencer did not detect these 2 mutations, another Ion Torrent run was performed on the same 7 samples for the 2 fragments that carried the mutation identified by Sanger sequence; fragment 2 and fragment 3. This was performed using a bi-directional method with a reverse barcode primer in addition to the forward barcode primer in order rule out strand-specific effects. The HbE mutation was detected by bi-directional sequencing but ambiguous results still remained in the 3 cases. Further investigation showed that the reason for these ambiguous results was that the primers used were not *HBB* specific and that the *HBD* gene was co-amplified generating a mixed sequence. The mutations found in the *HBB* gene are summarised in (Table 12.4).

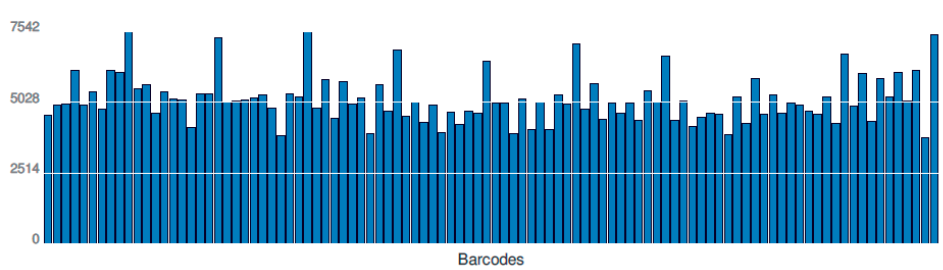

Figure 12.3. Coverage output of the evenly distributed sequence covering the first 100 samples.

Table 12.3. Summary of the artefacts obtained by the Ion Torrent sequence software due to the co-amplification of *HBB* and *HBD* fragments, showing the corresponding position in the *HBB* and *HBD* genes. These artefacts reside within fragment 3 of the *HBB* gene.

Artefact given by the Ion Torrent softaware in the HBB gene	Corresponding *HBB* position and nucleotide	Corresponding *HBD* position and nucleotide
c.[93-30G>C]	IVS-I-101, C	IVS-I-101, G
c.[93-28C>G]	IVS-I-103, G	IVS-I-103, C
c.93-24_93-23delAT	IVS-I-107 - 108, AT	IVS-I-107 - 108, TG (AT is deleted)
c.[93-19A>C]	IVS-I-112,T	IVS-I-112,T
c.[93-16T>C]	IVS-I-115,A	IVS-I-115,T
c.[93-9G>A]	IVS-I-122, C	IVS-I-122, C
c.[93-3A>G]	IVS-I-128, T	IVS-I-128, G
c.[93C>T]	Cd30, G	Cd30, A
c.[94G>A]	Cd31, C	Cd31, T
c.[96C>T]	Cd31, G	Cd31, A
c.[151T>A]	Cd50, A	Cd50, T

Table 12.4. List of mutations found in the Omani population by the ion torrent PGM sequencer. (HO = Homozygous, HR =Heterozygous).

Genotype	β-mutation HGVS name	no. of patients
Cd6 GAG>GTG (HO)	HBB:c.20A>T	85
IVS-I-5 G>C (HR)	HBB:c.92+5G>C	80
Cd6 GAG>GTG/IVS-I-5 G>C	HBB:c.20A>T/HBB:c.92+5G>C	26
Cd121 GAA>CAA (HR)	HBB:c.364G>C	9
Cd44 (-C) (HR)	HBB:c.135delC	9
IVS-I-1 G>A (HR)	HBB:c.92+1G>A	6
IVS-II-1 G>A (HR)	HBB:c.315+1G>A	6
IVS-I-3' (-25bp del) (HR)	HBB:c.93-21_96del	5
Cd39 CAG>TAG (HR)	HBB:c.118C>T	5
IVS-I-5 G>C (HO)	HBB:c.92+5G>C	4
Cd6 GAG>GTG/Cd44 (-C)	HBB:c.20A>T/HBB:c.135delC	4
Cd6 GAG>GTG/Cd121 GAA>CAA	HBB:c.20A>T/HBB:c.364G>C	4
Cd6 GAG>GTG/IVS-I-1 G>A	HBB:c.20A>T/HBB:c.92+1G>A	3
Cd6 GAG>GTG/Cd39 CAG>TAG	HBB:c.20A>T/HBB:c.118C>T	3
Cd5 (-CT) (HO)	HBB:c.17_18delCT	3
Cd30 AGG>ACG (HR)	HBB:c.92G>C	3
Cd26 GAG>AAG (HR)	HBB:c.79G>A	2
Cd44 (-C) (HO)	HBB:c.135delC	2

Table 12.4. List of mutations found in the Omani population by the ion torrent PGM sequencer. (HO = Homozygous, HR =Heterozygous). (*Continued*)

Genotype	β-mutation HGVS name	no. of patients
Cd6 GAG>GTG/IVS-II-1 G>A	HBB:c.20A>T/HBB:c.315+1G>A	2
Cd5 (-CT) (HR)	HBB:c.17_18delCT	2
Cd8 (-AA) (HR)	HBB:c.25_26delAA	2
5′ (-88) C>A (HR)	HBB:c.-138C>A	2
Cd6 GAG>GTG/Cd26 GAG>AAG	HBB:c.20A>T/HBB:c.79G>A	1
IVS-II-1 G>A (HO)	HBB:c.315+1G>A	1
IVS-I-5 G>C/IVS-I-3′ (-25bp del)	HBB:c.92+5G>C/HBB:c.93-21_96del	1
IVS-I-128 T>G (HR)	HBB:c.93-3T>G	1
Cd5 (-CT)/3′(+113) A>G	HBB:c.17_18delCT/HBB:c.+113A>G	1
3′(+108) - 3′(+112) 5nt del (HR)	HBB:c.+108_+112delAATAA	1
Cd26 GAG>AAG/IVS-I-3′ (-25bp del)	HBB:c.79G>A/HBB:c.93-21_96del	1
Cd129 TGC>TGT (HR)	HBB:c.389C>T	1
Cd22 GAA>TAA (HR)	HBB:c.67G>T	1
IVS-I-5 G>C/IVS-I-128 T>G	HBB:c.92+5G>C/HBB:c.93-3T>G	1
IVS-I-6 T>C (HR)	HBB:c.92+6T>C	1
Cd6 GAG>GTG/Cd121 GAA>AAA	HBB:c.20A>T/HBB:c.364G>A	1
Cd6 GAG>GTG/Cd6 GAG>AAG	HBB:c.20A>T/HBB:c.19G>A	1
Cd6 GAG>GTG/Cd5 (-CT)	HBB:c.20A>T/HBB:c.17_18delCT	1
Cd6 GAG>GTG/IVS-I-128 T>G	HBB:c.20A>T/HBB:c.93-3T>G	1
Cd121 GAA>CAA/IVS-I-1 G>A	HBB:c.364G>C/HBB:c.92+1G>A	1
Cd(8/9) +G (HR)	HBB:c.27_28insG	1
Normal controls		13
Total		297

DISCUSSION

The purpose of this study was to investigate if Ion Torrent PGM can be a suitable diagnostic method to sequence large number of samples for a small sized gene (e.g *HBB*) for a country wide Hemoglobinopathy screening program. Samples were preselected to contain 1 or 2 mutations in the beta globin gene by haematological and HPLC screening. Our primer design almost covered the complete *HBB* mutation spectrum worldwide except for a region in intron 2 as mutations in this region are very rare in the general population, at least not present in the Arabian area and not in the studied Omani population.

The reason for the observed artefacts in our samples and the discrepancy obtained in few cases was due to the co-amplification of the *HBD* gene in fragment 3. Existing sequence homology co-amplification is not always easily prevented. However, it should not cause a problem when data analysis is performed carefully and when necessary, Sanger sequencing

can be used as a confirmatory test. Moreover, although we have shown that a uni-directional method, by using a single barcode primer in the forward direction, is sufficient to give a unique sequence for each sample, it is always recommended to use bi-directional sequencing to discriminate between strand-specific artefacts and real mutations.

Based upon our results, we believe that Ion Torrent PGM could be a convenient option for diagnostic screening. One advantage of the PGM is its speed (8) and its ability to prepare a low-cost library with many different sample tags, allowing large number of samples to be pooled together and thus dividing sequencing costs among that many samples, generating sufficient data in a short period of time and avoiding lengthy hours of waiting to run samples in several batches. Moreover, PGM requires much less maintenance when compared to Sanger sequencing (4). Another advantage is that it uses a sequencing strategy where hydrogen ions (H^+) are detected (8), thus, no lasers, cameras or fluorescent dyes are needed, making it a cost-reasonable machine to purchase. Also, ion torrent's sequencing chips match the common manufacturing standards used for commercial microchips, meaning that it can be used for other technologies at equal low cost (9). Moreover, Ion Torrent requires as little as 10ng/ul of DNA to generate accurate results for challenging samples (10) whereas the traditional method of Sanger sequencing requires higher amount of DNA. Alternatively buccal swaps may be used instead of blood in a large scale screening programs. Although we have not tested the use of buccal swap DNA on Ion Torrent, we do believe that it should not be a problem for rapid population screenings. On the other hand, using buccal swap means having the DNA sample to be directly sequenced without haematological tests (CBC, HPLC) may save time and costs but will not provide with supportive data.

We have shown that the Ion Torrent PGM can be used to sequence relatively small genes (e.g *HBB*) of a large number of specimens rather than its traditional practice of sequencing highly complex genes on a small number of patients. However, before introducing Ion Torrent PGM sequence as a new diagnostic tool, few technical parameters should be focused at. First, manual PCR plate preparation raises chances of error; secondly, preparation of PCR plates is a lengthy procedure when multiple fragments are to be looked at. However, this could be avoided by the possibility of performing a multiplex PCR, amplifying the 8 fragments in a single PCR reaction in our case, thus, having 8x less PCR work. The reason why we amplified the fragments separately in this study was to ensure even coverage of every fragment since we used for the first time 10 x 10 combined barcodes to sequence 100 patients per run to have a robust and controllable amplification. To make PGM more feasible, more barcodes can be used to run as many patients (up to 400 samples) per chip. The reason why we have used a uni-directional sequencing in the first place was to reduce the number of primers and herewith the costs. However, when strand-specific sequence artefacts are expected, it is optimal to sequence bi-directionally. Economically speaking, cost will always remain an issue to see how worthy the new instrument can be in routine molecular analysis. However, increasing the sample size from 100 to 400 per run will reduce the costs per sample considerably and as the price per base continues to drop, PGM will soon be applicable routinely.

In conclusion, phenotypes cannot be predicted, explained or treated without determining the genotype. Our results show that the Ion Torrent Personal Genome apparatus may allow

genome sequencing in specialized reference labs and could be used within the screening program in Oman. With the new technology of PGM, multiple DNA sequence reads can be produced in a single run, generating sufficient data to solve the clinical and epidemiologic problems in a short time. Technically, the key to success would be to choose a methodological approach that avoids massive uneven PCR efficiencies and biased quantification and thus sequencing errors. A limitation of the ion torrent is the errors that arise from homo-polymeric stretches (4) that are being under- or over- called. However, *HBB* is very suitable to be analyzed by Ion Torrent PGM as it does not have homopolymers interfering with proper base-calling. Considering all that, Ion torrent could be a suitable technology for a national population screening program in Oman or elsewhere once DNA testing for HBP (or other diseases) becomes mandatory prior to marriage, promoting a quick tool of identification (diagnosis) and thus early prevention.

ACKNOWLEDGEMENTS

The authors declare to have conducted this study according to local ethical regulations and to have no conflicts of interest on the presented matters. Acknowledges goes to the Sardinian Regional Government for the financial support of P.O.R. Sadegna F.S.E. Operational Programme of Autonomous of Sardinia, European Social Fund 2007-2013 - Axis IV Human Resources, Objective I.3, Line of Activity I.3.1.

REFERENCES

1. Metzker ML. Sequencing technologies-the next generation. Nat Rev Genet 2010;11: 31-46.
2. Pourmand N, Karhanek M, Persson HH, et al. Direct electrical detection of DNA synthesis. Proc Natl Acad Sci USA 2006;103: 6466-6470.
3. Rothberg JM, Hinz W, Rearick TM, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;475:348-352.
4. Elliott AM, Radecki J, Moghis B, Li X, and Kammesheid At. Rapid Detection of the ACMG/ACOG-Recommended 23 *CFTR* Disease-Causing Mutations Using Ion Torrent Semiconductor Sequencing. Journal of Biomolecular Techniques. 2012; 23:24–30.
5. Knapp M, Stiller M and Meyer M. Generating barcoded libraries for multiplex high-throughput sequencing. Methods Mol Biol, 2012;840:155-170.
6. Dahl F, Stenberg J, Fredriksson S, Welch K, Zhang M, Nilsson M, Bicknell D, Bodmer WF, Davis RW and Ji H. Multigene amplification and massively parallel sequencing for cancer mutation discovery. Proc Natl Acad Sci. 2007;104:9387-9392.
7. Rozen S and Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 2000 132: 365-386.
8. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J and Pallen MJ. Performance comparison of benchtop high-throughput sequencing platforms. Nature biotechnology. 2012;30(5): 434-439.
9. Bolotin DA. Next generation sequencing for TCR repertoire profiling: platform-specific features and correction algorithms. Eur J Immunol. 2012; 42(11): 3073-3083.
10. Duhaime MB, Deng L, Poulos BT and Sullivan MB. Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method. Society for Applied Microbiology and Blackwell Publishing. Environmental Microbiology. 2012;14(9): 2526-2537.
11. Hassan SM, Hamza N, Al-Lawatiya F, Jaffer AM, Harteveld CL, Rajab A,Giordano PC. Extended molecular spectrum of beta- and alpha-thalassemia in Oman. Hemoglobin. 2010;34(2):127-34.