THE NEW MILLENNIUM

Application of advanced molecular technology in the diagnosis and management of genetic disorders in South Africa

M J Kotze, PhD

Division of Chemical Pathology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University and the National Health Laboratory Service, Tygerberg Hospital, Cape Town, South Africa

Corresponding author: M J Kotze (maritha@sun.ac.za)

Background. Genetic testing has evolved from a niche speciality for diagnosis of rare disorders and carrier screening to subtyping of complex medical conditions for targeted treatment. Genes causing monogenic disorders are well characterised, but risk management of multifactorial and polygenic disorders guided from the genetic background remains a challenge.
Objective. This study describes the use of a pathology-supported genetic testing (PSGT) strategy designed to facilitate the move from single- to multi-gene testing and next-generation sequencing (NGS).
Methods. In contrast to direct-to-consumer genetic testing, PSGT requires preselection of patients and data integration to determine current and future risk implications. To enable this process, a genomics database resource generated at the interface between the laboratory and clinic is available for clinical interpretation.
Results. The PSGT approach led to the development of testing algorithms for improved clinical management of patients with cancer and other complex disorders with a genetic component. Local evidence is presented to demonstrate the application of PSGT for assessment of clinical relevance in patients with rare germline variants and functional polymorphisms underlying shared disease pathways.
Conclusion. PSGT is ideally suited to serve as a screening step for microarray analysis and whole genome/exome sequencing as the next frontier in personalised medicine. Use of these advanced molecular technologies to match genotype with phenotype provides a resource for diagnosis and discovery over a lifetime.

S Afr Med J 2016;106(6 Suppl 1):S114-S118. DOI:10.7196/SAMJ.2016.v106i6.11012

Significant advances in molecular technology expanded the scope of genetic testing, from diagnostic confirmation and carrier screening of mendelian disorders, to risk management of complex polygenic and multifactorial disorders. A review by Pembrey and Beighton[1] on the molecular techniques that were available 30 years ago witnessed the transition from single- to multi-gene testing and next-generation sequencing (NGS) as arguably the most significant technological advancement in genetics to date.

The decreasing cost of microarray analysis and NGS, coupled with the use of sophisticated computational algorithms, enables genetic subtyping of complex disorders and identification of therapeutic targets at the same time. Whole genome/exome sequencing (WGS/WES) can now be performed alongside clinical evaluation as an integral part of personalised medicine.[2] This new healthcare model may be particularly useful when multiple aetiologies cause the same medical disorder, when the same disorder has multiple outcomes, and when response to treatment is unpredictable based on the

disease signs and symptoms alone.[3] However, lack of well-defined pre-screen algorithms and the ability to interpret complex genomic information are important barriers to clinical implementation. To bridge this gap, a pathology-supported genetic testing (PSGT) service was developed at the interface between the laboratory and clinical practice.[4] This enabled the establishment of a genomics database (freely available to registered users at www.gknowmix.org) for clinical interpretation and comparative effectiveness studies. Within a family context, PSGT combines medical history, environmental factors and pathology test results for interpretation of genetic findings across different assay platforms. Providing a patient report, as contextualised in Table 1, helps to determine whether any abnormal pathology or medication side-effects/failure may be caused by genetic or environmental risk factors, or both.

In this study, the incorporation of genomics into an existing body of knowledge is described, to provide a framework for NGS beyond the limited scope of single-gene testing. The application

Table 1. Matrix for data integration and semi-automated report generation using PSGT to identify risk factors underlying different disease subtypes in complex disorders

Disease pathway analysis	Family medical history and genetic susceptibility	Environmental factors and treatment response
Clinical risk profile	Contribution of genetic variants implicated in the dysfunctional regulation of key metabolic pathways across the disease spectrum to clinical presentation	Consideration of lifestyle intervention that may ameliorate risk for expression of disease-associated phenotypes in genetically susceptible individuals
Pathology test results	Pathological indicators (biochemistry, histology) which may reflect gene-environment interactions as biological intermediates	Monitoring of relevant pathological indicators/biochemical test results in relation to treatment response and side-effect profile

of PSGT in this context is explained in relation to examples of medical conditions, founder mutations, functional single nucleotide polymorphisms (SNPs) and pathology test results found to be suitable for inclusion or consideration as part of a pre-screen step before embarking on WGS/WES.

Single-gene testing of founder mutations and SNPs

Knowledge of the genetic structure of a population is very important for the development of diagnostic and screening tests. A founder gene effect may occur in some historically isolated populations that descended from a small gene pool. This phenomenon may account for the high frequency of certain cardiovascular, neurological, metabolic and intestinal conditions reported in South Africa (SA).[5] While disease penetrance is virtually 100% in the case of Huntington's disease, most genetic disorders show variable clinical expression. In patients with familial hypercholesterolaemia (FH) and variegate porphyria (VP), for example, clinical variability complicates accurate diagnosis of these autosomal dominant diseases, even in family members with the same founder mutation. Cascade screening is recommended for mutation-positive cases to improve diagnostic reliability in relation to gene-diet (nutrigenetics) and gene-drug (pharmacogenetics) interaction that may affect the same disease pathways.

Elucidation of the spectrum of high-penetrance mutations which cause FH in the majority of affected SA patients led to the development of a reverse-hybridisation strip assay for cardiovascular risk management.[6] Testing of the mutation panel underlying the high prevalence of FH in SA caused by a founder effect,[7] recurrent mutational events,[8] and population admixture,[9] may also be combined with apolipoprotein E (*APOE*) genotyping, to help distinguish between polygenic and monogenic FH.[10] PSGT provides an ideal framework for *APOE* genotyping in dyslipidaemics with a positive family history of late-onset Alzheimer's disease, incorporated as a novel component of the pre-screen algorithm developed by Lückhoff *et al.*[11] Physical activity was shown to attenuate the cholesterol-raising effect of the *APOE* e-4 allele, recognised as a poor target for lipid-lowering pharmacotherapy in international studies. This SNP occurs in 30 - 40% of the general population and affects lipid levels in a similar manner across ethnic groups.

While early detection, accurate diagnosis and optimal treatment of FH are priorities in SA due to the strong link with premature heart attacks, prevention of deep vein thrombosis (DVT) is equally important. Beighton and Richards[12] were the first to identify prolonged immobility during long-distance air travel as a potential environmental trigger for DVTs in individuals at increased risk for cardiovascular disease (CVD). In 1994, the factor V *Leiden* mutation (*1691G>A*) was identified as the most common genetic risk factor for DVT, causing a 7- to 80-fold increased risk for thrombosis in the presence of one or two copies of the faulty gene, respectively. Co-inheritance with the prothrombin mutation (*20210G>A*) found to increase the risk for venous thrombosis at least 40-fold, provided one of the first examples of an epistatic effect. While DVT is a relatively rare event in the general population (~1 in 1 000), both the factor II *20210G>A* (2 - 3%) and factor V *1691G>A* (3 - 5%) mutations occur in more than 1% of the Caucasian population. These SNPs are virtually absent in Africans and Asians, a finding that may contribute to differences in disease patterns of thrombophilia-related conditions among world populations. Identification of the target group most likely to benefit from SNP genotyping furthermore relates to the wide spectrum of environmental triggers, biochemical abnormalities (e.g. dyslipidaemia, hyperhomocysteinaemia) and medical conditions

(e.g. pulmonary embolism, recurrent pregnancy loss) associated with abnormal blood clotting. The factor V *Leiden* mutation is also a pharmacogenetic marker that can be used to recommend avoidance of certain forms of hormone therapy as well as smoking and obesity, in both heterozygotes and homozygotes.[13]

From single- to multi-gene testing

In 2004, the World Health Organization defined international standards for genetic testing, and a DNA test for the factor V *Leiden* mutation was the first to be approved for clinical use under these guidelines. At this important time in the history of human genetics, ethical concerns about genetic testing in SA were addressed in relation to a code of conduct compiled for the life insurance industry.[14] Analysis of thrombotic risk factors, individually or as part of a CVD multi-gene assay,[10] were evaluated in relation to: *(i)* the prevalence and penetrance of gene variants included in the test panel; and *(ii)* the potential benefits of preventative steps based on the findings. As a general rule, genetic testing will not have a negative effect when performed on the basis of well-defined selection criteria.

Owing to the increased risk of thrombophilia and cardiac complications with use of chemotherapy and hormonal therapy, the CVD multi-gene assay may also be applicable in breast cancer patients. Bardia *et al.*[15] reported that the predicted 10-year risk for CVD equals or exceeds breast cancer recurrence risk and urged physicians to offer early screening and cardiac prevention strategies for cancer survivors. CVD risk factor screening in 164 SA breast cancer patients showed a significantly higher body mass index (BMI) in oestrogen receptor (ER)-positive v. ER-negative cases, after adjustment for age, ethnicity and family history of cancer ($p<0.05$).[16] ER status and obesity were also linked to genetic variation in the folate-homocysteine methylation pathway, which provided valuable insight into the interplay between germline mutations and tumour pathology.

Understanding the molecular characteristics of both tumour and host genetics is critical to establishing their relationship with drug response and epigenetic processes underlying the development of cancer and many other chronic diseases. Nearly 100 genes have been identified that, if mutated, will convert a normal breast cell into a breast cancer cell. The influence of germline mutations on tumour pathology is particularly strong between mutations in the *BRCA1* gene and the basal-type breast cancer. This subtype usually tests negative for ER, progesterone receptor (PR) and human epidermal growth factor receptor-2 (HER2), and is, therefore, called triple-negative breast cancer (TNBC). ER, PR and HER2 status provide useful parameters for selection of patients eligible for transcriptional gene profiling, as evidenced in SA breast cancer patients referred for microarray analysis. Pohl *et al.*[17] demonstrated a change in chemotherapy treatment in 52% of SA patients with early-stage breast cancer by using a newly developed microarray pre-screen algorithm to facilitate risk assessment beyond standard pathology and clinical prediction tools.

While detection of germline mutations in the *BRCA1/2* genes is associated with a high risk for local or contralateral recurrence of breast cancer, microarray-based assessment of tumour genetics determines risk of distant recurrence (70-gene profile); and simultaneously enables subtyping of breast cancer into four treatment groups (80-gene profile): Luminal A, Luminal B, HER2-enriched and the basal-type. Owing to the ability of microarrays to distinguish between HER2-positive breast cancer of the Luminal B and HER2-enriched subtypes, our testing algorithm has now been extended to help resolve equivocal, borderline and contradictory pathology results prior to selection of patients for trastuzumab therapy. The

70-gene MammaPrint test has level A1 evidence for clinical utility and was cleared by the Food and Drug Administration (FDA) for use on fresh tumour biopsies (2007) and formalin-fixed paraffin-embedded (FFPE) tissue (2015).

Conventional medicine aims to diagnose and treat an existing disorder, while the purpose of personalised medicine is to predict the outcome of complex disorders and prevent them from recurring or developing in the first place. PSGT aims to integrate both medical models into a single genomics application. This approach has proven useful as a screening step to overcome some of the limitations of WES previously reviewed.[4]

From PSGT to exome sequencing

WES provides the opportunity not only to look into exclusive disease-specific genetic alterations, but also common molecular pathogenic mechanisms. However, poor coverage of repetitive or GC-rich gene regions and exclusion of promoters and other regulatory signals that may be located in introns are important limitations. To evaluate the depth of coverage in the case of *CYP2D6* genotyping considered applicable across clinical entities,[16] we compared WES with the results obtained using real-time polymerase chain reaction (PCR) technology. Table 2 shows the evaluation of twelve SNPs in four unrelated individuals. These include the index case (sample 4) of a family with three medical conditions previously analysed at the clinical, pathology and genetic level:[18] multiple sclerosis, hereditary haemochromatosis, and porphyria that may be triggered by iron dysregulation or drugs affecting CYP2D6 activity. Inadequate coverage (<20x) of the most common variant *CYP2D6*4* was evident in this patient (poor metaboliser), as well as samples 2 (intermediate metaboliser) and 3 (extensive metaboliser). Results obtained for this allele using the Taqman Drug Metabolism Genotyping Assay (Thermo Scientific, USA) were compared with WES in Table 2 for the four samples tested. Quantitative real-time PCR (qRT-PCR) revealed heterozygosity for *CYP2D6*5*, which results in hemizygosity at the *CYP2D6* locus and explains the apparent homozygosity for the variants defining alleles *4 and *10 in case 4. For heterozygous

Table 2. CYP2D6 genotypes analysed by WES, with coverage given below genotypes (CYP2D6 alleles listed according to position on chromosome 22)

Allele name	dbSNP ID	Ref	MAF[†]	Enzyme activity[‡]	Case 1	Case 2	Case 3	Case 4
CYP2D6*41	rs28371725	C	0.0934	Decreased	CC	CC	CC	CC
					119	88	69	128
CYP2D6*7	rs5030867	T	0	Absent	TT	TT	TT	TT
					91	91	38	42
CYP2D6*2	rs16947	G	0.343	Decreased	GG	AG	AG	GG
					90	82 (41/41)	34 (22/12)	31
CYP2D6*9	rs5030656	Ins	0.0258	Decreased	Ins	Ins	Ins	Ins
					91	61	23	11
CYP2D6*3	rs4986774	T	0.0254	Absent	TT	TT	TT	TT
					70	42	41	38
CYP2D6*4	rs3892097	C	0.186	Absent	TT[§]	CT[§]	CC[§]	TT[§]
					TT[¶]	TT[¶]	CC[¶]	CT[¶]
					23	3	15	18 (4/14)
CYP2D6*5	deletion	Ins	0.04	Absent	Ins[‖]	Ins[‖]	Ins[‖]	Ins/del[‖]
					-	-	-	-
CYP2D6*8	rs5030865	C	0	Absent	CC	CC	CC	CC
					60	45	43	22
CYP2D6*6	rs5030655	A	0.0199	Absent	AA	AA	AA	AA
					63	46	37	18
CYP2D6*29	rs61736512	C	0	Decreased	CC	CC	CC	CC
					65	40	32	20
CYP2D6*17	rs28371706	G	0.002	Decreased	GG	GG	GG	GG
					141	223	41	64
CYP2D6*10	rs1065852	G	0.202	Decreased	AA	AG	GG	AA
					263	113 (49/64)	107	322
Phenotype					PM	IM	EM	PM

Ref = reference allele; EM = extensive metaboliser; IM = intermediate metaboliser; PM = poor metaboliser.
[†] Minor allele frequency (MAF) according to dbSNP.
[‡] Enzyme activity according to Rebsamen *et al.*[20]
[§] *CYP2D6* rs3892097 genotype using the Taqman Drug Metabolism Genotyping Assay.
[¶] *CYP2D6* rs3892097 genotype using WES.
[‖] *CYP2D6* deletion using quantitative real-time PCR (qRT-PCR).

genotypes shown in Table 2, the numbers in brackets indicate the allele count of each alternate base at the variant position, in alphabetical order. The combined use of PCR and WES proved useful to resolve ambiguous results.

Case 3 was the only patient who tested negative for all 12 *CYP2D6* polymorphisms included in the initial gene panel, as well as the remaining part of the *CYP2D6* gene based on the WES results. Since this patient was diagnosed with both multiple sclerosis and TNBC breast cancer, *BRCA* mutation screening was also performed using WES. The result is shown in Fig. 1, indicating homozygosity for the microRNA 638 target SNP rs799917, previously implicated in TNBC progression via BRCA1 deregulation of DNA repair mechanisms.[19] The coverage for WES at this locus was sufficient to confidently identify this functional SNP as a true variant, as confirmed by extended genotyping in more than 100 cases and controls (>65 years) selected from our genomics database.

These findings contributed to improvement and analytical validation of the WES methodology used in our laboratory and confirmed the value of PSGT applied across different assay platforms. Early successes using the protocol outlined in Fig. 2 include the following findings based on adequate WES coverage and/or confirmation using Sanger sequencing (A V Peeters and M J Kotze, unpublished data):

• the detection of a novel 1-bp deletion in the *PALB2* gene in a patient with familial breast cancer previously shown to have a *BRCA1* variant of uncertain clinical significance;
• the detection of a missed 2-bp *BRCA2* deletion in a metastatic breast cancer patient who also tested positive for the *CYP2D6*4* allele;
• identification of a possible polygenic cause for FH in an SA family that tested negative for high-penetrance mutations in four known FH genes (*LDLR, APOB, PCSK9, LDLRAP1*).

WES reports were provided to the treating clinicians for family genetic counselling as appropriate. Cascade screening is indicated for the protein-truncating *PALB2* and *BRCA2* family-specific mutations. Low-density lipoprotein (LDL) cholesterol screening and regular monitoring are appropriate in cases with polygenic FH. Therapeutic recommendations based on *CYP2D6* activity are provided according to algorithms developed by the Dutch Pharmacogenetics Working Group.[13]
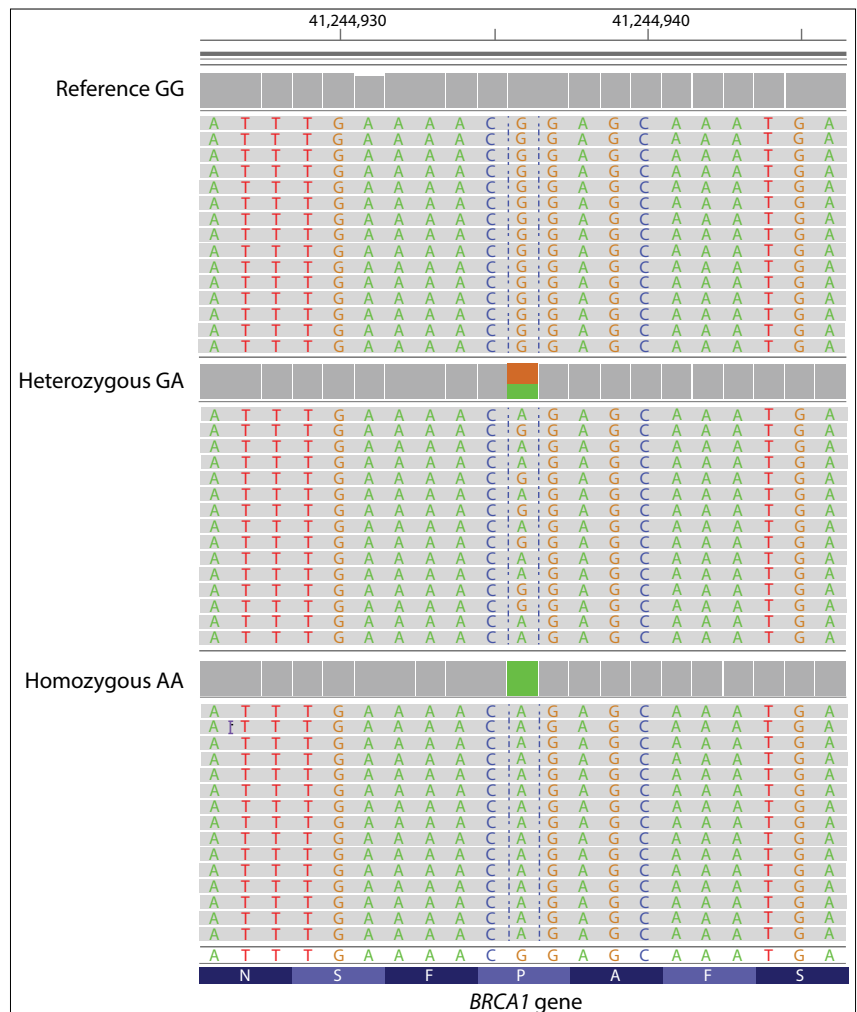

Fig. 1. Alignment view of whole exome sequencing reads encompassing the miR-638 target SNP rs799917 in the BRCA1 *gene.*
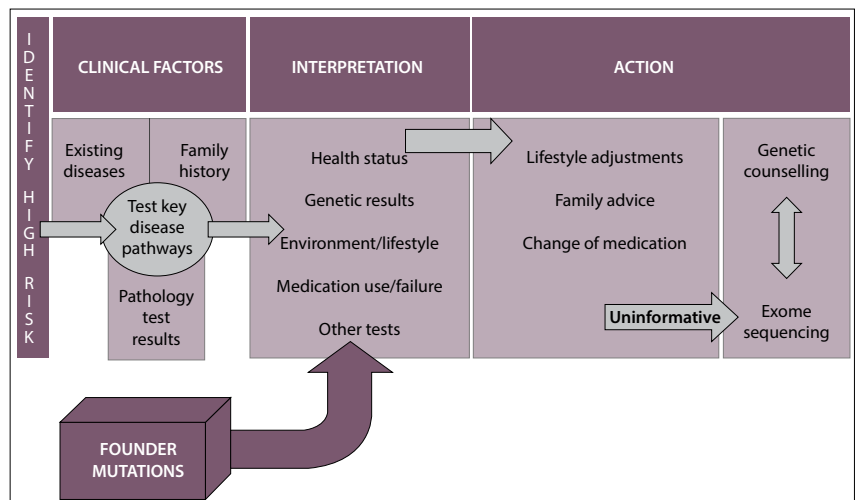

Fig. 2. Testing algorithm using pathology-supported genetic testing (PSGT) as a pre-screen step to determine eligibility for whole exome sequencing (WES). Focus areas for pathway analysis are: (i) lipid and lipoprotein metabolism; (ii) DNA methylation and mismatch repair; (iii) haemostasis and inherited thrombophilia; (iv) haem synthesis and iron homeostasis; and (v) drug metabolism.

Ethical considerations

Careful consideration is necessary before genetic testing is performed, since detection of a genetic predisposition in one individual implies that other family members may have the same mutation. Clinical interpretation

of genome scale data is complicated by the large number of gene variants of uncertain clinical significance uncovered by NGS, which may lead to unnecessary family screening. The human genome contains ~20 000 genes, with a considerable number of damaging mutations observed in healthy people that may never be expressed at the clinical level. Therefore, managing patient expectation regarding the information obtainable with WES/WGS requires knowledge of known gene-gene and gene-environment interaction.

According to the written informed consent provided by patients referred for PSGT, WES results are provided in the following circumstances, unless instructed otherwise:
• presence of a particular disorder or family history relevant to the test results
• detection of a genetic predisposition for a treatable or preventable condition
• genetic counselling required (e.g. family screening advisable irrespective of treatment implications).

Patients who consider WGS/WES need to be informed beforehand which genes or disease pathways will be analysed, or need to agree to return of results based on the three above-mentioned scenarios. Given the fast pace of genomic discoveries and the likelihood that new health concerns may present themselves over time, a once-off report may not be sufficient. If the raw sequencing data are stored securely or provided to the tested individual in electronic format together with the NGS report, future use will be possible for diagnostic and research purposes, according to the informed consent obtained at referral. A WGS/WES report that contains information on a limited gene panel only, without the option for further bioinformatics to be performed if necessary in future, would deny patients the full benefit of the test.

Conclusion
Referral guidelines and treatment options are well-defined for monogenic disorders, but for more complex disorders, PSGT applied in a learning healthcare system is ideally suited as a screening step for WES (Fig. 2). PSGT takes advantage of a founder gene effect in SA by combining analysis of a limited number of population-specific mutations with functional SNPs implicated in many chronic diseases, based on their role in drug response and cumulative risk. Understanding the genetics behind the onset and development of cancer and other complex disorders is critical for targeted drug treatment, which may act on the same metabolic pathways that caused cellular dysfunction in the first place. For this reason, a combined research and service delivery platform was established to cope with the demand for clinical interpretation of ever-increasing genomic information, while striving to meet patient expectations.

References
1. Pembrey M, Beighton P. Molecular technology in clinical genetics. S Afr Med J 1983;64(1):16-18.
2. Patel CJ, Sivadas A, Tabassum R, et al. Whole genome sequencing in support of wellness and health maintenance. Genome Med 2013;5(6):58. DOI:10.1186/gm462
3. Shelton CA, Whitcomb DC. Evolving roles for physicians and genetic counselors in managing complex genetic disorders. Clin Transl Gastroenterol 2015;2(6):e124. DOI:10.1038/ctg.2015.46
4. Kotze MJ, Lückhoff HK, Peeters AV, et al. Genomic medicine and risk prediction across the disease spectrum. Crit Rev Clin Lab Sci 2015;19:1-18. DOI:10.3109/10408363.2014.997930
5. Botha MC, Beighton P. Inherited disorders in the Afrikaner population of southern Africa. Part I. Historical and demographic background, cardiovascular, neurological, metabolic and intestinal conditions. S Afr Med J 1983;64(16):609-612.
6. Kotze MJ, Kriegshäuser G, Thiart R, et al. Simultaneous detection of multiple familial hypercholesterolaemia mutations facilitates an improved diagnostic service in South African patients at high risk of cardiovascular disease. Mol Diagn 2003;7(3-4):169-174. DOI:10.1007/bf03260034
7. Kotze MJ, Langenhoven E, Warnich L, du Plessis L, Retief AE. The molecular basis and diagnosis of familial hypercholesterolaemia in South African Afrikaners. Ann Hum Genet 1991;55:115-121. DOI:10.1111/j.1469-1809.1991.tb00404.x
8. Kotze MJ, Loubser O, Thiart R, et al. CpG hotspot mutations at the LDL receptor locus are a frequent cause of familial hypercholesterolemia among South African Indians. Clin Genet 1997;51:394-398. DOI:10.1111/j.1399-0004.1997.tb02497.x
9. Loubser O, Marais AD, Kotze MJ, et al. Founder mutations in the LDL receptor gene contribute significantly to the familial hypercholesterolemia phenotype in the indigenous South African population of mixed ancestry. Clin Genet 1999;55(5):340-345. DOI:10.1034/j.1399-0004.1999.550507.x
10. Kotze MJ, van Rensburg SJ. Pathology supported genetic testing and treatment of cardiovascular disease in middle age for prevention of Alzheimer's disease. Metab Brain Dis 2012;27(3):255-266. DOI:10.1007/s11011-012-9296-8
11. Lückhoff HK, Brand T, van Velden DP, et al. Clinical relevance of apolipoprotein E genotyping based on a family history of Alzheimer's disease. Curr Alzheimer Res 2015;12(3):210-217. DOI:10.2174/1567205012666150302154354
12. Beighton PH, Richards PR. Cardiovascular disease in air travellers. Br Heart J 1968;30(3):367-372. DOI:10.1136/hrt.30.3.367
13. Swen JJ, Nijenhuis M, de Boer A, et al. Pharmacogenetics: From bench to byte – an update of guidelines. Clin Pharmacol Ther 2011;89(5):662-673. DOI:10.1038/clpt.2011.34
14. Kotze MJ, Schörn D, Coetzer P. The impact of genetic testing on life insurance. J Genomics Afr Soc 2004;1:1-11.
15. Bardia A, Arieas ET, Zhang Z, et al. Comparison of breast cancer recurrence risk and cardiovascular disease incidence risk among postmenopausal women with breast cancer. Breast Cancer Res Treat 2012;131(3):907-914. DOI:10.1007/s10549-011-1843-1.
16. Van der Merwe N. Development and application of a pathology supported pharmacogenetic test for improved clinical management of South African patients with breast cancer and associated co-morbidities. PhD Thesis, Stellenbosch University. http://hdl.handle.net/10019.1/98481.
17. Pohl EH, Kotze MJ, Grant KA, et al. Impact of MammaPrint on clinical decision-making in South African patients with early stage breast cancer. Breast J 2016;April. DOI: 10.1111/tbj.12605. [Epub ahead of print]
18. Kotze MJ, de Villiers JN, Warnich L, et al. Lack of clinical manifestation of hereditary haemochromatosis in South African patients with multiple sclerosis. Metab Brain Dis 2006;21:109-120. DOI:10.1007/s11011-006-9015-4
19. Tan X, Peng J, Fu Y, et al. miR-638 mediated regulation of BRCA1 affects DNA repair and sensitivity to UV and cisplatin in triple-negative breast cancer. Breast Cancer Res 2014;16(5):435. DOI:10.1186/s13058-014-0435-5
20. Rebsamen MC, Desmeules J, Daali Y, et al. The AmpliChip CYP450 test: Cytochrome P450 2D6 genotype assessment and phenotype prediction. Pharmacogenomics J 2009;9(1):34-41. DOI:10.1038/tpj.2008.7