Genomic medicine and risk prediction across the disease spectrum

Maritha J. Kotze1, Hilmar K. Lückhoff1, Armand V. Peeters1, Karin Baatjes2, Mardelle Schoeman3, Lize van der Merwe3,4, Kathleen A. Grant1,5, Leslie R. Fisher1, Nicole van der Merwe1, Jacobus Pretorius1, David P. van Velden1, Ettienne J. Myburgh5, Fredriek M. pienaar5, Susan J. van Rensburg8, Yandiswa Y. Yako8, Alison V. September9, Kelebogile E. Moremi8, Frans J. Cronje10, Nicki Tiffin11, Christianne S. H. Bouwens12, Juanita Bezuidenhout1, Justus P. Apffelstaedt2, F. Stephen Hough12, Rajiv T. Erasmus7, and Johann W. Schneider1

1Division of Anatomical Pathology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa, 2Department of Surgery, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa, 3Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University and Tygerberg Hospital, Cape Town, Western Cape, South Africa, 4Department of Statistics, University of the Western Cape, Cape Town, Western Cape, South Africa, 5Department of Biomedical Sciences, Faculty of Health and Wellness, Cape Peninsula University of Technology, Cape Town, South Africa, 6Panorama Medi-Clinic, Cape Town, South Africa, 7GVI Oncology, Panorama Medi-Clinic, Cape Town, South Africa, 8Division of Chemical Pathology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University and National Health Laboratory Service, Cape Town, South Africa, 9UCT/MRC Research Unit for Exercise Science and Sports Medicine, Department of Human Biology, University of Cape Town, Cape Town, South Africa, 10Department of Interdisciplinary Medicine, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa, 11South African National Bioinformatics Institute/Medical Research Council of South Africa Bioinformatics Unit, University of the Western Cape, South Africa, and 12Department of Internal Medicine, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

Abstract

Genomic medicine is based on the knowledge that virtually every medical condition, disease susceptibility or response to treatment is caused, regulated or influenced by genes. Genetic testing may therefore add value across the disease spectrum, ranging from single-gene disorders with a Mendelian inheritance pattern to complex multi-factorial diseases. The critical factors for genomic risk prediction are to determine: (1) where the genomic footprint of a particular susceptibility or dysfunction resides within this continuum, and (2) to what extent the genetic determinants are modified by environmental exposures. Regarding the small subset of highly penetrant monogenic disorders, a positive family history and early disease onset are mostly sufficient to determine the appropriateness of genetic testing in the index case and to inform pre-symptomatic diagnosis in at-risk family members. In more prevalent polygenic non-communicable diseases (NCDs), the use of appropriate eligibility criteria is required to ensure a balance between benefit and risk. An additional screening step may therefore be necessary to identify individuals most likely to benefit from genetic testing. This need provided the stimulus for the development of a pathology-supported genetic testing (PSGT) service as a new model for the translational implementation of genomic medicine in clinical practice. PSGT is linked to the establishment of a research database proven to be an invaluable resource for the validation of novel and previously described gene-disease associations replicated in the South African population for a broad range of NCDs associated with increased cardio-metabolic risk. The clinical importance of inquiry concerning family history in determining eligibility for personalized genotyping was supported beyond its current limited role in diagnosing or screening for monogenic subtypes of NCDs. With the recent introduction of advanced microarray-based breast cancer subtyping, genetic testing has extended beyond the genome of the host to also include tumor gene expression profiling for chemotherapy selection. The decreasing cost of next generation sequencing over recent years, together with improvement of both laboratory and computational protocols, enables the mapping of rare genetic disorders and discovery of shared genetic risk factors as novel therapeutic targets across diagnostic boundaries. This article reviews the challenges, successes, increasing inter-disciplinary integration and evolving strategies for extending PSGT towards exome and whole genome sequencing (WGS) within a dynamic framework. Specific points of overlap are highlighted

Keywords
Breast cancer, database, exome, genome, hg19, major allele reference sequence, metabolic syndrome, multiple sclerosis

History
Received 6 July 2014
Revised 3 November 2014
Accepted 20 November 2014
Published online 19 January 2015
between the application of PSGT and exome or WGS, as the next logical step in genetically uncharacterized patients for whom a particular disease pattern and/or therapeutic failure are not adequately accounted for during the PSGT pre-screen. Discrepancies between different next generation sequencing platforms and low concordance among variant-calling pipelines caution against offering exome or WGS as a stand-alone diagnostic approach. The public reference human genome sequence (hg19) contains minor alleles at more than 1 million loci and variant calling using an advanced major allele reference genome sequence is crucial to ensure data integrity. Understanding that genomic risk prediction is not deterministic but rather probabilistic provides the opportunity for disease prevention and targeted treatment in a way that is unique to each individual patient.

Abbreviations: AOL: Adjuntiv! Online; APOE: apolipoprotein E; BMI: body mass index; BRCA1: breast cancer gene 1; BRCA2: breast cancer gene 2; BIG: Breast International Group; CYP2D6: Cytochrome P450, family 2, subfamily D, polypeptide 6; CVD: cardiovascular disease; DTC: direct-to-consumer; EDSS: expanded disability status scale; ER: estrogen receptor; FH: familial hypercholesterolaemia; FISH: fluorescence in situ hybridization; FFPE: formalin fixed paraffin embedded; GWAS: genome-wide association studies; HH: hereditary hemochromatosis; HER2: human epidermal growth factor receptor 2; IHC: immunohistochemistry; ITM: intima media thickness; MDD: major depressive disorder; MTHFR: methylene-tetrahydrofolate reductase; NCDs: non-communicable diseases; NAFLD: non-alcoholic fatty liver disease; NASH: non-alcoholic steatohepatitis; PSGT: pathology-supported genetic testing; PCR: polymerase chain reaction; PR: progesterone receptor; RCTs: randomized controlled clinical trials; RASTER: microarray prognostics in breast cancer; MS: multiple sclerosis; RT-PCR: reverse transcriptase polymerase chain reaction; SNPs: single nucleotide polymorphisms; TMPRSS6: transmembrane protease, serine 6; TNBC: triple-negative breast cancer; VaD: vascular dementia; WES: whole exome sequencing; WGS: whole genome sequencing

Introduction

A dramatic epidemiological transition in the prevalence from acute to chronic non-communicable diseases (NCDs) has been observed over the last century. Major technological advances in genomics are well placed to support a new medical model that may be required in this context, in keeping with the ideals of personalized medicine. However, with the exception of a restricted number of established clinical applications, personalized genomics, recognized as a form of medicine that uses information about genes, proteins and the environment for improved risk management, has yet to be embraced to the extent that was initially anticipated following completion of sequencing of the human genome in 2003.

During most of the history of genetics, it has been a purely scientific field, limited to the laboratory. Genetic testing gained a small clinical role as a diagnostic tool for a number of hereditary or congenital disorders, but with limited therapeutic implications. As in the case of Huntington’s disease, in spite of the ability to identify the gene defect, the fact that no therapeutic options exist resulted in reluctance of many patients to undergo pre-symptomatic genetic testing. Therapeutic implications of genetic diseases started entering clinical practice in two areas: first, in the realm of risk reduction therapy based on the identification of deleterious germline mutations in at-risk individuals. Familial colon and breast cancer are two examples where testing of DNA extracted from blood, saliva or buccal swabs for germline mutations allowed for surgical intervention to reduce the risk of developing cancer. Second, the search for a cure for cancer based on tumor behavior led to the development of several biological agents targeting the kinase pathway in the cell cycle. Genetic tests such as fluorescence in situ hybridization (FISH) used in combination with immunohistochemistry (IHC) became the gold standard for detection of human epidermal growth factor receptor-2 (HER2) over-expressing breast cancer using formalin fixed paraffin embedded (FFPE) tissue. Although response rates can be dramatically improved by anti-HER2 treatment in a subgroup of patients, predictive markers for chemotherapy selection only became available with the more recent introduction of transcriptional profiling for breast cancer prognostication. Based on these developments, a multi-disciplinary model of cancer risk assessment and clinical management was established, which now provides a solid foundation for extended incorporation of genomics in clinical practice.

In spite of more than two decades of clinical application of genetics, most clinicians have little training and experience in using genetic tests to guide therapeutic decision making. There seems to be a progressively widening gap between the explosion in novel genomic information and the ability to apply this new knowledge to clinical patient care. In contrast to the initial expectation of a “one gene, one disease” concept, genetic aspects of NCDs are complicated by the variable penetrance of disease-causing mutations and disease-associated polymorphisms that may interact with various biological, physical and social factors. It has therefore become important to develop enrollment guidelines for participation in genomics-based risk assessment and management programs formulated from evidence supporting the performance and clinical utility of ever-increasing number of genomic tests. Against this background, a pathology-supported genetic testing (PSGT) service was developed to identify subgroups of patients requiring different treatment strategies. An integrated service and research approach is used for establishing and populating a research database to facilitate the following:

(2) Evaluation of the performance of new genomic tests against the gold standard for improved quality assurance (analytical validation).

(3) Documenting treatment decision making and health outcomes over time to determine the impact of genomics in a clinical setting outside a controlled research environment (clinical utility).

The need for a new clinical model to facilitate responsible communication of complex genomic risk information, led to the establishment of a central research database in parallel with routine genetic testing service delivery using an ethically approved protocol (project reference numbers N09/06/166 and 09/08/224, Stellenbosch University). Following informed consent from the patient, a questionnaire is used to capture data on personal and family medical conditions, medication use and side effects, lifestyle factors and pathology test results relevant to the genetic analysis performed. A waiver of consent may also apply to specific ethically approved projects where potential benefit to society is considered greater than the perceived risk that may be related to such a sub-study. The central genomics database resource (accessible to registered users at www.gknowmix.org) was developed to facilitate research translation at the interface between the laboratory and clinic.

While randomized controlled clinical trials (RCTs) have been the standard for developing new therapeutic interventions, they remain costly to design and run, require large data sets and need lengthy follow-up to assess the final outcome. In many cases, the development of new technologies outpaces the reporting of trial results. Alternative models are being employed to allow faster data acquisition and earlier adoption of new treatment options. One example of this approach is the use of neo-adjuvant trials in oncology, where the effect of chemotherapy can be seen on the tumor tissue as a result of the intervention before removal. Similar substitute models have been suggested for evaluation of complex genomic applications. When aiming to match a biomarker profile with drug treatment, the impact of the genomic test on clinical decision-making is an important determinant of the level of regulation required. The use of prospectively collected patient data and tissue registries have recently gained acceptance as an alternative to RCTs, allowing careful monitoring of clinical outcome in relation to the genomic test results. Such surrogate trials involving retrospective evaluation of randomized, prospectively collected cohorts may allow conditional approval of new genomic applications, provided that the selection criteria for the data set conform to the disease of interest.

The use of DNA testing to identify high-penetrance mutations causing monogenic disorders is well-evidenced. Thousands of susceptibility variants identified by genome-wide association studies (GWAS) are also implicated as potential risk modifiers in both monogenic and complex disease traits. While of limited or no value in risk prediction when used in isolation, there is growing recognition of the clinical usefulness for actionable functional gene variants underlying key metabolic pathways across diagnostic boundaries. In this context, sophisticated bioinformatics tools are increasingly applied to identify clinically relevant genes acting on the same biological pathway. The discovery of novel high impact rare variants and copy number repeats facilitated by next generation whole exome sequencing (WES) and whole genome sequencing (WGS) lends further support to the notion that sporadic polygenic NCDs are in many respects more similar to Mendelian disorders than previously anticipated. An arbitrary distinction between the role of causative mutations in monogenic disorders versus rare variants and single nucleotide polymorphisms (SNPs) in polygenic diseases therefore seems superficial and belies the underlying genetic complexity characteristic of disease pathogenesis in general. Recognition of this genetic intricacy highlighted the need to define an optimal strategy by which a convergent pathogenic model, incorporating clinically relevant genomic knowledge generated at the DNA, RNA and/or protein levels, may be integrated into public health research, policy and practice.

In this review, we describe the development of the clinically integrative PSGT approach to genomics-based NCD risk assessment and management as an enabler of research translation. The need for standardized referral guidelines and eligibility criteria for genetic testing is first discussed in the context of breast cancer genomics as a health discipline where this requirement has been recognized and successfully addressed. We consider the example of how assessment of the underlying tumor pathology prompted re-evaluation of existing guidelines for BRCA mutation analysis (germline susceptibility testing) and gene expression profiling (tumor pathology) for chemotherapy selection beyond the limited scope of identifying monogenic breast cancer subtypes. Establishment of a database resource provided the foundation for development of test selection criteria for genomics-based NCD risk management programs, focused on shared disease pathways across seemingly unrelated illnesses. Insights gained as a result are guiding the development of optimal strategies, whereby PSGT itself could be positioned as a pre-screening tool to facilitate selection of patients eligible for next-generation sequencing. Combination of these technologies may facilitate genome-wide studies in search of susceptibility alleles for cardio-metabolic risk, as well as the identification of genetic drivers within cancer genomes. Our initial focus on the exome is justified by cost efficiencies, as the coding sequence and intron-exon splice junctions are currently the most informative regions of the genome for clinical application. Moreover, the research component linked to the PSGT service provides a valuable platform for validation of new genomic applications against current standards, given the critical importance for ongoing comparison with non-sequenced based techniques, including bioinformatics tools increasingly applied in clinical care.

Pathology-supported genetic testing across the disease spectrum

Insight gained from extensive genetic research performed in the diverse South African population led to development of the PSGT concept, whereby pathology test results are combined with genetic testing in an attempt to identify a genetic component to the disease phenotype, while at the same time providing information on relevant future risk implications. This clinically integrated model applied as part of a combined open-innovation research and service delivery
platform, is positioned as a solution to the inherent limitations of direct-to-consumer (DTC) genetic testing. DTC genetic testing is based on an incomplete view of the lifetime risk for common complex diseases, as this model does not include simultaneous evaluation of the personal and family medical history relevant to the tests performed. PSGT involves a biological pathway systems approach to clinically orientated genomics as a means of guiding patient management within the context of underlying pathology as vantage point, as illustrated in Figure 1.

PSGT uses a multidisciplinary framework for the interpretation of genetic information with the goal of identifying treatable NCD subtypes requiring a more active approach to therapeutic management. A top-down (phenotype-to-genotype) and bottom-up (genotype-to-phenotype) approach can be applied simultaneously for consideration of patient data obtained from diverse fields of healthcare linked via the common thread of pathology. PSGT aims to empower the clinician with genomic information that can be used to facilitate (1) the diagnosis of monogenic/treatable disease subtypes, (2) lowering of cumulative risk for disease progression or the development of co-morbidities and (3) formulation of individualized treatment strategies tailored to the needs of the individual. Moreover, the development of tailored pharmacogenomic algorithms aimed at predicting treatment failure and limiting exposure to serious drug side-effects may ultimately serve to improve treatment compliance. A clinically guided genomics-based approach to NCD risk assessment may therefore address current as well as future risk, with the ultimate goal of optimizing health and patient management throughout the course of an individual’s lifespan.

Rooted in extensive research performed in South Africa concerning genetic underpinnings of both monogenic and polygenic diseases, PSGT aims to facilitate evaluation of risk factors and choice of intervention strategy in patients with etiologically complex multifactorial disorders. Application of this dynamic healthcare model was first described using biochemical abnormalities as the vantage point to facilitate improved risk management in genetically pre-disposed patients with abnormal iron status or high cholesterol levels. Detection of a genetic contribution to iron overload enables confirmation of a diagnosis of hereditary hemochromatosis (HH) without the need for an invasive liver biopsy, while accurate diagnosis of familial hypercholesterolemia (FH) is required to determine the appropriateness of long-term use of cholesterol-lowering medication in dyslipidemic patients. Identification of founder mutations underlying the high prevalence of these conditions in high-risk population groups led to development of cost-effective single-gene tests that may also be applied as part of multiplex genomic applications targeting critical disease pathways of relevance to a wide spectrum of chronic NCDs. While the clinical utility of single-gene testing for the spectrum of mutations defined as causative in a high-risk population group is well-established, evaluation of multiple low-penetrance mutations/SNPs found to be of relevance across diagnostic boundaries is complicated by uncertainty related to selection criteria and interpretation of the results for clinical application. PSGT aims to overcome these limitations and may also serve as a screening step for selection of patients eligible for WES/WGS. This approach is similar to the strategy used routinely to identify patients with a family history and/or clinical features suggestive of familial breast cancer for full gene BRCA screening (Table 1), following the exclusion of known causative mutations underlying the majority of affected cases due to a founder effect. In genetically uncharacterized patients where a particular disease pattern or therapeutic intolerance/failure is not adequately accounted for, extension from PSGT to WES/WGS could further be considered for the identification of rare or novel causative mutations or pathogenic pathways not covered as
Table 1. Application of pathology-supported genetic testing (PSGT) in breast cancer risk management.

<table>
<thead>
<tr>
<th>Process</th>
<th>Evaluation</th>
<th>Important considerations prior to genetic testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document clinical risk profile, relevant pathology data and family history in the context of genetic counseling support</td>
<td>Personal risk</td>
<td>Early age at diagnosis (&lt;40 years). Presence of bilateral breast cancer and/or ovarian cancer.</td>
</tr>
<tr>
<td></td>
<td>Ancestral risk</td>
<td>A limited number of founder mutations may account for the disease in the majority of affected patients in high-risk populations (e.g. Ashkenazi Jewish, Afrikaners).</td>
</tr>
<tr>
<td></td>
<td>Familial risk</td>
<td>Breast/ovarian cancer runs in the family from one generation to the next. In addition, other cancers including colon, prostate and male breast cancer are important indicators especially if at least four close relatives younger than 60 years are affected, three below 50 years or two below 60 years and ovarian cancer in the family.</td>
</tr>
<tr>
<td></td>
<td>Histopathology</td>
<td>Triple-negative breast cancer (ER-, PR-, HER2-negative) and patients with basal-like subtype are at increased risk to carry a BRCA1 mutation. Receptor status important for referral of patients for transcriptional profiling used in anti-HER2 treatment and selection of chemotherapy.</td>
</tr>
<tr>
<td></td>
<td>Treatment failure/side-effects</td>
<td>Use of certain antidepressants may influence Tamoxifen response due to reduced CYP2D6 activity, found to be of particular concern in BRCA mutation carriers. CYP2D6 metabolizes about 25% of commonly used prescription drugs and is therefore of relevance in a wide range of clinical domains.</td>
</tr>
<tr>
<td>Define BRCA1/2 testing approach</td>
<td>First screen an affected family member to identify the causative mutation</td>
<td>Determine high familial risk for recurrence or bilateral breast cancer and increased risk of ovarian/other cancer in the context of treatment options.</td>
</tr>
<tr>
<td></td>
<td>Screen close relatives for known family mutation to exclude or confirm the familial risk</td>
<td>Determine risk to inherit a BRCA mutation and based on the outcome the likelihood to develop cancer over the lifetime in the context of treatment options.</td>
</tr>
<tr>
<td>Pre-test genetic counseling</td>
<td>Value of the test in a family member without cancer</td>
<td>A positive test would indicate a very high risk for cancer that requires intensified surveillance or prophylactic surgery with proven efficacy. A negative test for the family mutation reduces the familial risk to that in the general population.</td>
</tr>
<tr>
<td></td>
<td>Value of the test in a patient already diagnosed with cancer</td>
<td>A positive test indicates a high risk for recurrence/bilateral breast cancer or ovarian cancer. A negative test result for the family mutation means the cancer was caused by other risk factors not tested for (extended mutation screening may be recommended based on overall risk profile).</td>
</tr>
<tr>
<td></td>
<td>Risk to offspring</td>
<td>A 50% chance with each pregnancy to pass on the faulty BRCA gene (mutation-positive individuals). No further risk implications for children when the causative family mutation was excluded.</td>
</tr>
<tr>
<td>Decide when to test</td>
<td>Above the age of 18 years</td>
<td>Women who need to have a bilateral oophorectomy usually prefer to know their risk before the operation. Women who consider breast surgery may use the information on BRCA mutation status as motivation for a bilateral mastectomy. Allows family planning and adequate screening to be timed between pregnancies. Decision about timing of risk reduction surgery based on genetic risk and expression of disease occurring earlier in subsequent generations.</td>
</tr>
<tr>
<td>Choose BRCA test option</td>
<td>Mutation-specific</td>
<td>Screen at-risk family members for a known mutation previously identified in the index patient diagnosed with breast/ovarian cancer.</td>
</tr>
<tr>
<td></td>
<td>Population-specific</td>
<td>Screen for a limited number of mutations occurring at an increased frequency in cancer patients from certain population groups (e.g. Ashkenazi Jews, Afrikaners of European descent).</td>
</tr>
<tr>
<td></td>
<td>Full gene screen</td>
<td>Screen for the cancer-causing mutation in breast cancer patients who tested negative during the initial screen for family/population-specific mutations or when these two test options are not indicated.</td>
</tr>
<tr>
<td>Provide option for research participation centered around the storage, decoding or destroying</td>
<td>CYP2D6 pharmacogenomics testing as part of a chronic disease screening program aimed at lowering of cumulative risk in breast cancer survivors</td>
<td>Ongoing studies enable the research team to improve the tests and interpretation of results as new discoveries and actionable information become available in the scientific literature.</td>
</tr>
</tbody>
</table>
part of the PSGT pre-screen framework. In order to prepare for such change, it is important to note that the need for genetic counseling will increase exponentially as WES/WGS becomes a viable alternative to targeted sequencing in genetically uncharacterized patients.

**Determining eligibility for oncogenomic testing**

Breast cancer oncogenomics is currently leading the way in translating genetic discoveries into actionable clinical benefits. Since the role of genetic risk factors underlying different subtypes of breast cancer is well-evidenced and testing options have become part of clinical reality, research efforts have been actively pursuing cost-effective strategies to identify South African breast cancer patients who may benefit from DNA-based testing and/or RNA-based gene profiling for targeted treatment. Development of selection criteria for the increasing number of genomic applications was recognized as an important research area to identify where additional information, if any, provided by a genomic test could fit into the context of current clinicopathological prognostication of breast cancer.

Standardized selection criteria have been developed for inclusion of patients in familial breast cancer screening programs based on the high lifetime risk (40–85%) associated with the BRCA1 and BRCA2 tumor suppressor genes. BRCA mutations only account for 30–40% of familial disease clustering and less than 5% of total cases overall. Causative mutations therefore remain undetected in the majority of affected families with currently used screening methods, while detection of BRCA1/2 sequence variants of unclear pathogenic significance constitutes an increasing clinical challenge. RNA profiling has therefore become an attractive option to determine functionality of gene variants identified, and at the same time select patients sensitive to new therapeutics focused on the genetic defect identified. Global RNA profiling and promoter methylation analysis performed on 70 breast tumor biopsies also revealed the involvement of hereditary factors in non-BRCA1/2 breast cancer families. The finding that family members may carry genetic susceptibility not only to breast cancer but also to a particular subtype of breast cancer implies that RNA profiling can be used to identify functional subgroups within breast tumors, particularly in families that tested negative for BRCA1/2 germline mutations. In this context, genetic variation underlying dysfunction of the folate-homocysteine methylation pathway was shown to predict both the subtype of breast cancer and disease progression. HER2-enriched and basal-like subtypes were positively associated with familial breast cancer and inversely associated with plasma folate. The clinical relevance of an extensively studied multi-functional SNP in the methylene-tetrahydrofolate reductase (MTHFR) gene (rs1801133, 677 C>T) was supported by strong association with the Luminal B subtype. The Luminal A subtype was associated with lack of family history of breast cancer, late age of onset and higher body mass index (BMI). A growing body of evidence supported by WES suggests that rare variants and co-inheritance of multiple low-penetration mutations may explain the majority of familial disease clustering not attributable to BRCA1/2 mutations. Functional polymorphisms in genes coding for drug-metabolizing enzymes or those dependent on vitamin co-factors for optimal activity may explain the variable clinical expression and differences in treatment response, drug toxicity and/or recurrence risk in patients with the same diagnosis. Variation in the MTHFR and Cytochrome P450, family 2, subfamily D, polypeptide 6 (CYP2D6) genes are considered of particular importance in this context across diagnostic boundaries, due to their role in gene–diet (nutrigenomics) and gene–drug (pharmacogenomics) interactions. Appropriate clinical application related to these aspects will depend on whether the patient is at risk or has already been diagnosed with cancer and therefore correct interpretation of genomic information facilitated by the PSGT approach is a critical step in patient care.

It is now appreciated that, rather than constituting a single clinical diagnosis, breast cancer defines a broad spectrum of pathology accounting for significant inter-patient heterogeneity in presentation, prognosis and treatment outcomes. More accurate determination of breast cancer recurrence risk based on the underlying tumor biology has become possible with use of microarray analysis that enables disease classification into at least four intrinsic molecular subtypes. BRCA1 mutations were found to be more common in patients with triple-negative breast cancer (TNBC) that lacks expression of estrogen receptor (ER), progesterone receptor (PR) and HER2. Different breast cancer subtypes defined partly by these three biomarkers furthermore showed variable patterns of association with NCD’s, as indicated by the relationship between TNBC and the metabolic syndrome. The metabolic syndrome, characterized by central obesity, hypertension, dyslipidemia and impaired glucose tolerance, is considered an important unifying risk factor of shared disease.
pathways implicated in a wide spectrum of NCDs. The role of epigenetics in dysfunction of the methylation pathway discussed above is well-established and may also cause silencing of the BRCA1 gene in sporadic breast cancer. Recent findings that BRCA1 and BRCA2 mutations are more common in patient subgroups with sporadic breast cancer than previously anticipated have important implications for genetic testing. Variation in the CYP2D6 gene associated with variability in treatment response was found to be of particular relevance in BRCA mutation carriers, especially when concomitantly treated with antidepressants that may affect enzyme function. CYP2D6 genotyping is therefore considered appropriate in South African breast cancer patients who are receiving Tamoxifen treatment and are at high recurrence risk due to inheritance of a defective BRCA gene in the family and/or when using potential competing antidepressants. Van der Merwe et al. proposed that the same path used for development and implementation of genetic tests for high-penetration mutations such as those in the BRCA1/2 genes, should also apply to variants with reduced penetrance, with the exception that known environmental triggers documented as part of the PSGT approach should also be taken into account for clinical application.

The above-mentioned findings collectively highlight the role of genetic variation not only in familial, but also in sporadic breast cancer cases and recurrence risk. Use of specific test selection criteria is aimed at a high probability for detection of causative mutations in the index case and pre-symptomatic diagnosis in at-risk family members, based on the age of onset and a strong family history of breast, ovarian and other relevant cancer types previously shown to be associated with the BRCA1/2 genes. Since founder-related mutations in the BRCA1 and BRCA2 genes account for the majority of cases of monogenic breast cancer in the South African population and other high-risk groups such as the Ashkenazi-Jewish population, an individual’s ancestral origin is also an important criterion in determining eligibility for high-penetration BRCA mutation screening. Test options include DNA-based BRCA mutation analysis for diagnosis of familial breast cancer using (1) family-, (2) population-specific mutation detection methods, or (3) full gene screening of BRCA1 and BRCA2. Selection of any of these three test options during pre-test genetic counseling is based on previous identification of the causative mutation in an affected family member or population risk due to a founder effect, whereby a limited number of common causative mutations generally accounts for the disease in the majority of affected patients.

In addition to consideration of the above-mentioned factors focused primarily on selection of breast cancer patients for BRCA mutation screening, tumor morphology and receptor status are increasingly assessed to determine eligibility for gene expression profiling. Microarray analysis represents a higher level of investigative complexity and use of the PSGT approach was essential for successful introduction of breast cancer gene profiling in clinical practice, as evidenced by reimbursement of the 70-gene MammaPrint profile as part of the South African breast cancer gene expression profiling trials. Unlike BRCA gene testing, the MammaPrint test does not provide information about the risk for inheritance of cancer in the family, but whether a patient with early stage breast cancer may benefit from chemotherapy or not. The prognostic and predictive value of this microarray test was confirmed in both the adjuvant and neo-adjuvant settings, as further supported by prospective 5-year follow-up data. Despite the lack of supporting data from a RCT, this level of evidence was considered sufficient to allow conditional approval for routine clinical use of MammaPrint in South Africa. An important requirement was that a central database be established for (1) evaluation of predefined referral guidelines outside a controlled research environment, (2) evaluation of the performance of microarray technology against the gold standard for improved quality assurance and (3) to determine the impact on clinical decision making and clinical outcome. As discussed below, these aspects were adequately addressed and led to successful introduction of advanced microarray-based breast cancer gene profiling into the South African healthcare system.

Utilization of a genomics database resource for comparative effectiveness studies

The favorable impact of the 70-gene MammaPrint profile on long-term clinical outcome supports the value of patient registries, allowing for the evaluation of prospective follow-up data in comparative effectiveness studies as a viable alternative to RCTs. Advanced microarray technology has proven to be highly accurate in facilitating risk stratification and guiding not only chemotherapy selection in resource-limited settings, but also HER2-targeted treatment based on patient re-classification following reflex testing. This implementation, facilitated by utilization of the central research database, substantiated the clinical usefulness of the MammaPrint Pre-screen Algorithm (MPA), validated as an appropriate strategy to prevent chemotherapy overtreatment in South African patients with early-stage breast cancer. Risk assessment is provided over and above ER, PR and HER2 status (routinely performed in all breast cancer patients using IHC/FISH) as a separate read-out (TargetPrint) of the versatile MammaPrint microarray platform. From 2011, molecular subtyping into Luminal A, Luminal B, basal-like and HER2-enriched subtypes based on receptor pathway activity (BluePrint) was also added to the MammaPrint service.

Given the concerns over the accuracy of standard techniques used for determination of HER2 status in breast cancer patients, incorporation of RNA-based microarray analysis may facilitate clinical care by resolving borderline cases and providing a second opinion on standard protein- and DNA-based HER2 testing. While the accuracy of reverse transcriptase polymerase chain reaction (RT-PCR) has previously been questioned for determination of HER2 status, the TargetPrint results obtained in 138 tumor specimens of South African breast cancer patients showed 100% concordance between microarray-based and IHC/FISH results upon reflex testing in discordant cases. Repeat IHC/FISH testing confirmed the accuracy of TargetPrint for determination of HER2 status, irrespective of whether fresh or FFPE tumor specimens were used. FFPE tumor specimens were found to be a reliable source of RNA for microarray analysis of HER2 status, analytically validated in South African breast cancer patients using TargetPrint.
In contrast to centralized HER2 testing utilized in large international breast cancer trials, most tests used in clinical practice in South Africa are performed by different local laboratories. They all employ international quality control programs, but this does not necessarily equate to 100% accuracy in the day to day reports used in clinical decision making. Use of a central research database proved invaluable to address practical clinical issues experienced by clinicians and demonstrated that, in patients undergoing MammaPrint genomic profiling, TargetPrint provides a reliable ancillary method of assessing HER2 status in conjunction with IHC/FISH. Furthermore, addition of the 80-gene BluePrint profile as a separate read-out from the MammaPrint microarray platform allows prediction of therapeutic outcomes based on the identification of functional pathways implicated in carcinogenesis and neoplastic transformation, independent of hormone and HER2 receptor status. Microarray-based breast cancer gene profiling provides information not available through standard histopathology and enables quantification of discordant/equivocal cases for appropriate interpretation of results obtained at the DNA, RNA and protein levels.

The significant impact of the 70-gene MammaPrint test on treatment decisions in South African breast cancer patients has recently been confirmed by Pohl et al. in an external audit of the central genomics database. Although only those patients expected to gain little benefit from additional chemotherapy were selected for this audit, while still having the option of endocrine therapy, gene profiling could potentially change the treatment of one in two (52%) patients irrespective of 10-year mortality based on clinical risk assessment using the Adjuvant! Online (AOL) tool. Mortality figures determined by AOL relates to statistical comparison between groups and do not reflect the specific risk for an individual patient. Genomic profiling determines the programming or behavior of the tumor relative to a cut-off point between low and high risk for metastases, regardless of external morphology. Therefore, when a patient is classified as high risk using MammaPrint, the AOL risk becomes obsolete. In the prospective MicroRay Prognostics in Breast Cancer (RASTER) trial, a 5-year distant recurrence-free interval of 100% was reported in MammaPrint low-risk patients who did not receive adjuvant chemotherapy, despite the presence of “high-risk” clinicopathological factors. The risk of distant recurrence is approximately 30% in patients with a high-risk MammaPrint profile, while low risk patients have a 10% risk of recurrence within 10 years. This translates into a three times higher risk of dying in high risk versus low risk patients. When chemotherapy is added there would be no significant reduction in cancer deaths in the MammaPrint low-risk patients, while the high risk group would benefit from approximately 25% risk reduction.

Establishment of a central genomics database also allowed the identification of patients referred for a specific genetic service who might be at risk of other hereditary or chronic NCDs. Accordingly, some of the patients referred for microarray analysis of their breast tumors participated in the chronic disease screening program and/or were tested for germline mutations in the BRCA1 and BRCA2 genes based on their clinical profile and familial risk. Preliminary data obtained from comparative studies suggest an association between BRCA1- and BRCA2-positive breast cancer and a high-risk MammaPrint profile, independent of ER status. Although pre-selection using the MPA in the majority of patients tested may partly account for this finding, it is important to note that tumors in BRCA1 positive patients are more frequently triple-negative in comparison to tumors in BRCA2 carriers. This supports previous findings showing that having a first-degree family history of breast cancer is associated with an increased risk of TNBC. Furthermore, increased incidences in TNBC and obesity may be related to shared disease mechanisms associated with the components of the metabolic syndrome. Cumulative risk due to clustering of metabolic syndrome features including central obesity, hypertension, dyslipidemia and/or glucose intolerance in the same patient, may lead to development of cardiovascular disease (CVD) found to exceed the risk of breast cancer recurrence at 10 years in breast cancer survivors. Thrombophilia or cardiotoxicity induced by some of the therapeutic options for breast cancer is also a major concern and warrants the development of pharmacogenomic treatment algorithms aimed at prevention of drug failure and side effects.

Modifiable environmental risk factors including alcohol consumption, smoking and obesity, shown to have a significant effect on the recurrence rate of breast cancer, should also be taken into account in relation to the impact of genetic variation in a high-risk environment.

BRCA mutations have been linked to an increased risk of various types of cancer including risk for development of lung cancer, found to at least double in smokers with the BRCA2 rs13314271 variant due to a combined effect. An interaction between the presence of BRCA mutations and diet diversity in relation to fruit and vegetable intake has also been reported. These findings support the implementation of preventative and therapeutic strategies based on our understanding of the mechanisms through which genetic and environmental factors contribute to cancer development, progression and recurrence risk. For application of personalized medicine to be clinically meaningful, genomic information needs to be considered together with clinical and environmental influences. This aspect is addressed by ongoing collection of clinical, pathological, lifestyle and genetic information in the PSGT research database, substantiated by the need for monitoring of breast cancer patients at risk of both physical and psychological complications.

The metabolic syndrome as a unifying risk factor for common NCDs

Obesity and associated conditions such as dyslipidemia, hypertension and type II diabetes are responsible for many preventable diseases and premature death. The use of a genomics-based approach targeting common disease mechanisms across the diagnostic spectrum may therefore be best exemplified by considering the constellation of cardio-metabolic risk factors (collectively termed the metabolic syndrome) as a unifying pathogenic model accounting for the development or progression of multiple NCDs (Table 2), in addition to breast cancer as discussed above.
The PSGT research database has proven to be an invaluable resource for validation of novel and previously described gene-disease associations replicated in the South African population for a broad range of clinically related NCDs associated with cardio-metabolic risk factors and features of the metabolic syndrome. The significant positive association observed between BMI and homocysteine levels, correlating with low folate intake in the diet, confirmed the deleterious effect of a genetic disturbance in the methylation pathway in patients with multiple sclerosis (MS) and depression. The effect of environmental factors known to influence BMI and homocysteine were evaluated in these studies not only as potential confounders that need to be adjusted for during statistical analysis, but also as potential modifiable contributors to the disease process that may in turn be useful to mitigate the gene effect. In MS patients, smoking was associated with disease progression, while daily intake of at least five fruits and vegetables correlated with a favorable MS disability score as measured by the Expanded Disability Status Scale (EDSS). Furthermore, a recent ultrasonography study investigating vascular risk factors in South African MS patients, found a highly significant correlation between intima media thickness (ITM – a surrogate marker for CVD) and disability in MS based on the EDSS. Evidence that vascular risk may be modified by lifestyle factors in a way that counteracts the effect on MS progression warrants a prospective implementation study to determine whether lifestyle intervention, possibly guided from the genetic background, will result in improved clinical outcome.

The shared bidirectional pathogenic relationship between the metabolic syndrome and neuropsychiatric illnesses including major depressive disorder (MDD) provides the rationale for routine evaluation of overall cardio-metabolic risk in all patients diagnosed with MDD. Evidence that elevated homocysteine levels in 86 MDD South Africans studied by Delport et al. may be mediated by a combined effect of the MTHFR 677C>T mutation and low folate intake on BMI, prompted extended assessment of MDD patients for presence of the metabolic syndrome. In this pilot, the majority of MDD patients were obese (BMI >30 kg/m²) and waist circumference measured in these patients confirmed central obesity as a defining characteristic of the metabolic syndrome (Figure 2). A 14.6% increase in BMI was noted for each additional feature of the metabolic syndrome present in the study cohort, for fixed age and gender (Figure 3). As previously described in South African patients at risk of CVD, the metabolic syndrome was defined by the presence of three or more of the following features: central obesity characterized by waist circumference >102 cm in males and >88 cm in females, blood pressure >130/85 mmHg, fasting glucose ≥5.6 mmol/L, HDL-cholesterol <1.0 mmol/L in males and <1.3 mmol/L in females and triglyceride levels ≥1.7 mmol/L. While alcohol intake was not associated with an increasing number of metabolic syndrome components identified (p = 0.460), a significant negative association was found with physical activity (p = 0.006). This finding reflects the importance of exercise to reduce the disease risk associated with the metabolic syndrome.

A multitude of varied factors have been proposed to explain the co-existence of MDD and the metabolic syndrome, including unhealthy diet and lifestyle habits, poor general access to healthcare, hypothalamic shift and the side-effects of medications such as tricyclic antidepressants and atypical antipsychotics, which can cause weight gain and
The metabolic syndrome is regulated by the presence of three or more of the following features: central obesity characterized by waist circumference ≥102 cm in males and ≥88 cm in females, blood pressure ≥130/85 mmHg, fasting glucose ≥5.6 mmol/L, HDL-cholesterol <1.0 mmol/L in males and <1.3 mmol/L in females and triglyceride levels ≥1.7 mmol/L.

Unfavorable metabolic changes. Emotional stress disrupts the integrity of the hypothalamic–pituitary–adrenal axis and increases the production of cortisol, noradrenaline and cytokines, considered of particular relevance in South African patients with a genetic predisposition for early onset MDD. Since the diagnosis of a life-threatening disease such as breast cancer may also increase the risk of depression in South African patients, assessment of BMI as a useful correlate of central obesity represents an important clinical indicator for participation in the chronic disease screening program.

Genetics of dyslipidemia in relation to chronic disease risk management

Each component of the metabolic syndrome is regulated by both genetic and acquired factors resulting in variable phenotypic expression. A genotype-only approach to NCD risk management therefore profoundly limits the capacity for the clinical translation of genomic research. Testing for well-evidenced and clinically actionable functional variants such as the apolipoprotein E (APOE) polymorphism included in the CVD multi-gene assay developed in South Africa is considered relevant to identify environmental and epistatic drivers of clinical expression in complex disease traits. Comprehensive patient assessment considering diverse medical, biochemical, genetic and lifestyle data is used to develop tailored intervention strategies guided partly from a genetic background to decrease cumulative NCD risk in genetically susceptible individuals.

The majority of South African patients who present biochemically with type III dysbetaIoproteinemia are homozygous for the epsilon-2 allele (ε-2) of APOE. The deleterious effect of this allele is triggered by a second hit such as obesity, diabetes and hypothyroidism. This is in accordance with an increasing number of metabolic syndrome features found to be associated with the APOE ε-2 allele in South African patients at risk of CVD. In this context, it is important to note that hepatic manifestation of the metabolic syndrome, termed the non-alcoholic fatty liver disease (NAFLD), is characterized by increased intra-hepatocellular accumulation of triglycerides in excess of 5% of liver mass. Several studies have shown that serum triglycerides, total and LDL cholesterol levels are increased, and HDL cholesterol levels decreased, in NAFLD patients. Similar but more severe derangements in lipid profiles are observed in patients with progressive hepatic necro-inflammation, which defines the more aggressive inflammatory steatohepatitis (NASH) phenotype. A high-fat, Western-style diet appears to unmask the association between APOE ε-2 and impaired lipid homeostasis, which could promote progression to NASH.

The aforementioned genotype–phenotype associations require careful consideration in the context of the metabolic syndrome and/or NAFLD, given the possibility of coexisting FH, due to an increased prevalence of this severe form of dyslipidemia in South Africa as a consequence of a founder effect. Confirming a suspected diagnosis of FH has profound implications for the clinical management of patients and at-risk family members, as it necessitates early initiation of long-term lipid-lowering pharmacotherapy to effectively reduce risk for adverse cardiac events as well as cognitive decline. In comparison, APOE ε-4 allele carriers without FH may not only be less responsive to both lipid-lowering statins and medications commonly used in the treatment of dementia, but also more susceptible to deleterious dietary and lifestyle habits associated with greater cumulative risk for CVD and dementia. In an investigation of the role of APOE ε-4 as a genetic risk factor for dyslipidemia in South Africans with and without FH, Kotze et al. found that the APOE ε-4 allele was associated with hypercholesterolemia in the unaffected control group, while an additive effect on the lipid profiles was not apparent in patients with FH. This finding and underrepresentation of the APOE ε-4 allele in genetically characterized FH patients studied, suggested that the genetic effect of this variant is masked or overridden by the known causative FH mutation.

APOE ε-4 is also a well-established genetic risk factor for late-onset Alzheimer’s disease and a major determinant of inter-patient heterogeneity, which influences nearly every pathogenic domain affected in this condition. Its hypercholesterolemic effects promote accelerated atherogenesis implicated not only in ischemic cerebrovascular disease associated with Alzheimer’s disease as well as vascular dementia, but also microvascular dysfunction underlying white matter pathology observed in subcortical dementias and certain Alzheimer’s disease subtypes. This raised the possibility that having a family history of Alzheimer’s disease may potentiate cumulative risk associated with phenotypic expression of the APOE ε-4 allele in an additive or synergistic manner. In order to determine whether expression of the known dyslipidemic effects attributable to polymorphic variation in the APOE gene are modified by a family history of Alzheimer’s disease, data from more than 500 South African individuals participating in the genomics-based chronic disease screening program were evaluated, based on the presence or absence of a known family
Alcohol consumption is known to potentiate the association between HDL cholesterol as previously demonstrated by Djousse revealed genotyping errors that led to a request for subgroup of asymptomatic hypercholesterolemics set to genotyping as a means of identifying a high-risk history of Alzheimer’s disease. In this study, we showed that

Genomic medicine and risk prediction across the disease spectrum

From genome-wide association studies to next generation clinical sequencing

GWAS identified a large number of genetic variants involved in disease susceptibility and drug response, but translation into clinical practice is slow due to lack of proven clinical utility for the majority of SNPs. While WES/WGS is considered an important advancement in overcoming the limitations of GWAS-generated genomic information, it is imperative that the strengths and weaknesses of next generation sequencing be objectively assessed.

The growing use of WES/WGS in the research context and clinical domain represents a robust means of detecting causative mutations of variable penetrance and high impact rare variants not possible before. However, the complex nature and vast scale of both the next-generation sequencing technologies and bioinformatics systems employed in WES/WGS-based identification of deleterious genetic variants may result in many problems. These range from library preparation and sequencing artifacts, to the data analysis pipelines employed to filter and annotate potential causative gene variants. A significant shortcoming of WES in particular is the fact that this sequencing methodology cannot identify disease-associated structural genetic abnormalities, trinucleotide repetitions and most intronic variants with their potential regulatory functions. Another significant limitation is the relatively high error rate for incorrect base pair identification. Genotype and variant inference in the presence of such misalignment and base-reading sequencing errors, without confirmation by alternative mutation detection methods, may therefore lead to the incorrect assignment of risk-allele carriage with potential downstream implications for patient management. Any candidate etiological variant detected by next generation sequencing analysis must be additionally confirmed by another laboratory technique such as Sanger sequencing to define the presence of the variant allele with confidence. Furthermore, the process of identifying candidate etiological variants from the many private mutations of unknown clinical significance detected in individuals and families, especially those of African origin given the vast pool of undocumented genetic diversity on the continent, is a complicated and time-consuming process. Given the fact that it is seldom possible to identify the etiological variant with absolute certainty without further functional studies, significant ethical challenges are posed in the return of especially novel findings with potential clinical implications to patients.

Clinical application of WES/WGS may further be complicated by lack of sufficient infrastructure and data storage systems compatible with the large amount of genetic...
information generated by next generation sequencing technologies. The development of enabling computational and bioinformatics tools are therefore required. Detailed phenotypic description and characterization of patients subjected to such testing are also required to optimize the use of next generation sequencing for risk assessment and management purposes at the individual level. This requirement is partly provided by using the PSGT approach that includes generation of genetic information on a limited number of founder mutations common in South Africa, where relevant and/or clinically useful functional polymorphisms implicated in disease pathways shared by many NCDs. For quality assurance purposes, the known gene variants covered in the chronic disease screening program as part of the PSGT service could be verified by WES/WGS prior to a search for disease causative mutations in genetically uncharacterized patients or non-responders to drug treatment. This step serves to confirm the identity of the patient sample analyzed, while newly identified gene variants in turn may be validated in an extended sample of patients and controls from the existing database resource as illustrated in Figure 1, in order to exclude genotype errors resulting from the use of WES/WGS and to validate novel findings in a clinical context.

While most WES studies published in 2013 were performed on the Illumina HiSeq, alternative sequencing platforms have emerged with the commercial release of the Personal Genome Machine (PGM). This apparatus is based on non-optical detection of hydrogen ions released with the sequential addition of deoxynucleotides to a growing DNA chain and may offer a faster turn-around time, less costly than current fluorescent-labeled detection platforms. Despite these advantages, the PGM apparently does not generate WES data with adequate high coverage per base. In 2012, the Proton was released, which is considered the “next generation” of semiconductor sequencing instrumentation, designed to generate gigabases of data at a low cost. Compared to the typical 6-day run time of the Illumina HiSeq, the Proton provides results within one day, with 3.5 h run time and 8 h data processing time under ideal conditions. Major discrepancies were identified with both platforms in the detection of small indels due to the propensity for polymerase slipping during polymerase chain reaction (PCR), resulting in sequencing artifacts. Since indels are expected to account for a large fraction of somatic changes in cancer, further refinement of the technical sequencing and calling algorithms used for sequencing data generated by both the Illumina and Proton is required. Boland et al. demonstrated that the Proton can generate high-quality WES data and detect a number of SNPs and indels not identified by the Illumina, which in turn identified variants that were missed by the Proton.

In our experience from previous results obtained from the first four exomes of MS patients sequenced on the Illumina HiSeq2000 next generation sequencing platform, followed by repeat sequencing in 2014 using the Proton as discussed below, more than one next generation sequencing platform should ideally be used to obtain high-quality data suitable for clinical application. In addition, the low concordance between variant-calling pipelines poses a significant potential source of error affecting the accuracy of WES/WGS sequencing data.

Exome sequencing in multiple sclerosis research: lessons learned

Extension of PSGT to WES was first applied in adult patients with MS to further explore the role of genetic risk factors mediating the apparent interplay between iron metabolism and inflammation in a subgroup of patients. In accordance with a recent study performed in more than 100 MS patients, two adult MS cases were heterozygous for a non-synonymous (V736A) change in the transmembrane protease, serine 6 (TMPRSS6) gene. This SNP (rs855791) previously identified by GWAS reduces the ability of the enzyme to inhibit the transcription of hepcidin, the main regulator of iron metabolism. Since inflammation, implicated in all NCDs and various other environmental and genetic factors, also influences hepcidin expression, TMPRSS6 rs855791 is included as a part of the PSGT gene panel that may be extended to WES as illustrated in Figure 1. Due to ethical constraints related to testing of children for genetic risk factors pre-disposing to adult-onset medical conditions, the standard PSGT pre-screen step was omitted in two South African children selected for WES, who were found to be highly responsive to iron supplementation. Homozygosity reported initially for TMPRSS6 A736V (rs855791) using WES could, however, not be confirmed afterwards by conventional Sanger sequencing in our laboratory. This discrepancy was caused by the presence of the risk-associated minor T-allele of TMPRSS6 rs855791 in the public reference genome sequence (hg19) used for variant calling, instead of the wild-type major C-allele. This finding also highlighted the fact that where a major allele is replaced by the minor allele in the reference genome sequence, true homozygotes will be missed by WES and most likely filtered out due to an apparent high population frequency. Fortunately, these problems were identified prior to return of relevant genetic findings to MS patients with low iron status.

The unexpected high number of false-positive variants initially reported in our patient cohort did not only affect relatively common SNPs as may be expected for TMPRSS6 rs855791, but also rare variants due to incorrect presentation of the major allele as the minor allele. The presence of minor alleles in hg19 may negatively affect both the sequence read alignment and variant calling used in WES. Individuals who are homozygous for a minor allele listed in the reference genome will not be identified as having a gene variant at that locus and the opportunity for interpretation of clinical relevance will be missed. Both heterozygotes and homozygotes for major alleles absent in hg19 would furthermore be reported as potential disease-associated variants at such loci. This supports the development of a catalog of clinically annotated variants (SNP knowledge database) as an essential element of the PSGT platform containing information extracted from the literature on the risk allele and trait/disease associations previously replicated in different populations. Results from meta-analyses are freely available online in databases such as the Single Nucleotide Polymorphism Database of the National Centre for Biotechnology Information (www.ncbi.nlm.nih.gov/SNP) and SNPedia that gathers information from PubMed on a daily basis to support genome annotation and interpretation (http://www.snpedia.com/index.php/SNPedia).
Data retrieved after variant calling are individually evaluated against these resources to verify their minor/risk allele status, prior to analytical validation of potentially deleterious gene variants in the laboratory.

Since a corrected reference genome sequence was not readily available to solve the variant calling problem following WES, we searched the scientific literature for the possible existence of a major allele reference sequence that can be used instead of hg19. Such a tool was indeed described by Dewey et al., who reported the critical importance of an enhanced human reference sequence for accurate disease-, carrier- and pharmacogenetic-risk assessment using WES/WGS. These authors reported 10,396 variant alleles using hg19 in a family study, compared to only 9,389 variant alleles identified when the synthetic ethnically concordant major allele reference genome was used for variant calling. The genotyping error rate was 38% higher with use of the hg19 reference genome (5.8 per 10,000 variants) compared with the major allele reference genome (4.2 per 10,000 variants). A total of 233 genotype calls at 130 disease-associated SNPs differed between the hg19 and the major allele reference genomes. Independent genotyping technology verified 161/188 genotypes (85.6%) in the major allele call set compared to 68/188 (36.2%) in the hg19 call set. Preliminary findings following variant calling in one of the above-mentioned South African MS patients revealed discordance of nearly 20% between use of hg19 and the major allele reference sequence.

Although not relevant in the case of de novo mutations, routine inclusion of both affected and unaffected family members when performing WES is important to exclude potential errors based on the fact that both parents of homozygous children would be obligate heterozygotes. Transmission analysis, physical pair-end sequencing of libraries and population-based genome phasing analysis performed in a family quartet studied by Lajugie et al. enabled the generation of a more complete list of true variants and determination of the false-positive and false-negative variant calling error rates in comparison with hg19. The hg19 reference genome was deduced from a collection of DNA samples from anonymous individuals, but when compared against a large high-quality SNP disease-association database, 3,556 disease-associated variants, including 15 rare variants, were found. Since the interpretation of personal genome sequences mainly focuses on the identification and functional annotation of genetic variants that differ from hg19, these findings highlighted the fact that the presence of minor alleles relating to disease-susceptibility in this public reference human genome should be taken into account during data interpretation for assessment of personal disease risk.

Next-generation sequencing presents a conundrum to the field of medical science: while it has the potential to identify every single causal or contributory gene variant for any given genetic disorder, every step along the path to this goal is complicated by technical, clinical and ethico-legal challenges. These difficulties range from phenotyping and informed consent for sample collection to discrepancies between different sequencing platforms and low concordance among multiple variant-calling pipelines used in WES/WGS.

Ethical considerations

We are rapidly approaching a reality when data interpretation capacity and ethical concerns rather than cost barriers will be the main limiting factors in genomic medicine. In the South African context, fear of genetic discrimination in relation to life insurance was identified as a major concern upon completion of the sequencing of the human genome in 2003. To address this issue, three areas of risk were defined at the time to explain that all genetic tests are not equal:

1. Medical conditions transmitted by a single dominant gene or a pair of recessive genes.
2. Pre-disposition to a disease or group of diseases with only a small proportion of the disease burden caused by single gene defects.
3. Course of disease is not genetic in nature but is influenced by genetic variation.

The benefits of genetic testing need to be weighed against potential harm and would depend on various factors, including cost, the prevalence and penetrance of a gene variant, the mortality associated with the mutation/disease and the potential for effective intervention based on the additional genetic knowledge gained. Both the emergence of WES/WGS and a growing proliferation of large-scale genetic biobanking ventures internationally have further complicated the ethical landscape of genomics and biomedical research. These developments magnified existing moral dilemmas inherent to genetic testing. For example, firmly held beliefs on what constitutes true voluntary informed consent and the process by which this should be obtained have effectively been challenged by the advent of large-scale biobanking. New population-based or context-dependent models may need to be developed in order to foster research in previously underrepresented African population groups in relation to WES/WGS research.

Data security and privacy are important concerns that need to be carefully considered in relation to the use of an open innovation platform allowing for sharing of genetic data. General topics to be considered include accountability, informed consent, transparency, disclosure and patient de-identification to ensure privacy. Ideally, the establishment of an effective data integration system requires (1) standardization of genetic information, (2) adequate infrastructure to facilitate analytical validation of assays and assist in the interpretation, reporting and follow-up of genetic data, (3) the development of effective archiving methods as well as (4) statistical tools to simplify appropriate analysis of data. To achieve these outcomes, a computer-based data integration system is continually being developed in our laboratory, linked to a secure centrally maintained research database and related software programs. In accordance with these goals, the PSGT platform enables the documentation of patient information collected at different professional levels, for integration in such a manner as to provide a comprehensive report with actionable information to the referring clinician.

While it is generally agreed that genetic data may be returned to individuals under the condition that they are clinically actionable, results generated in the context of WES/WGS range in clinical relevance and the extent to which research and service delivery may impact upon the degree and
timing of data return. Due to the unknown significance of a substantial amount of data generated by these technologies, the legal, and to a lesser extent, moral duty of the researcher to return such information is limited at this time. The duty to inform is further influenced by the relationship between researcher and participants and may vary depending on the setting, geographic and cultural background. Yu et al.\textsuperscript{128} suggested providing participants with the option of self-selection of specific results obtained via WES/WGS as a favorable alternative, which respects the principles of beneficence, non-maleficence and autonomy, as well as allowing for future access as deemed appropriate. The question also arises whether data sharing for secondary analysis implies retaining of responsibilities to return these results to participants. While it remains unclear how clinicians will disseminate the large volumes of genetic data generated by WES/WGS, the need for a greater role of genetic counselors as well as medical scientists participating in the service delivery process seems obvious. Pre-test counseling forms an integral part of genetic research and includes relaying information on the voluntary nature of the inclusion. However, given the sheer volume of information generated as a result of WES/WGS, it seems inconceivable that clinicians will be able to anticipate all the implications for their patients. For WES/WGS to be successfully incorporated as part of clinical practice, it is imperative that appropriate methods of providing pre-test counseling be developed, to guide informed decision making, and that clinicians are equipped to interpret and communicate such WES/WGS applications to patients.

Our experience in the ethical return of genetic testing results using a clinician-mediated combined service delivery and research approach provides a sound basis from which the ethico-legal challenges posed by next-generation sequencing can be addressed. Potential concerns are pre-emptively addressed by positioning the PSGT service as a pre-screening tool for (1) selection of patients eligible for WES/WGS, and (2) to extend the current gene panel towards a more comprehensive analysis of clinical and pathology results in the context of the genetic findings. The ultimate aim is to empower the clinician to derive the best current, clinically actionable treatment. Follow-up assessment of long-term health outcomes has long been considered primary indications of patient referral for genetic testing. In this overview, we explain how the consideration of pathological criteria (reflected by histological or biochemical assessment), alongside clinical risk factors incorporated in the development of a genomics pre-screen algorithm, is used to assist clinicians in determining patient eligibility for genomic testing. Insights gained as a result of research performed in the diverse South African population over more than two decades, provide a useful framework for the positioning of the PSGT service as a pre-screening tool to facilitate the selection of genetically uncharacterized patients or non-responders to treatment, eligible for, and set to benefit from WES/WGS. The addition of such eligibility guidelines may limit unnecessary patient referral for next-generation sequencing in cases where individual disease risk, pathological severity or adverse clinical or therapeutic outcomes, can be sufficiently explained by genetic variants incorporated as part of the carefully selected multi-gene testing assays utilized in the PSGT approach. Given the finding that the public reference genome sequence (hg19) contains minor alleles at more than 1 million positions\textsuperscript{121}, the correct allocation of the minor allele for each potential risk-associated variant identified after variant calling presents a critical step for appropriate interpretation of genetic data.

The MammaPrint experience substantiated the value of a genomics database as a viable alternative to RCTs for establishing the clinical utility of emerging genomic applications. A practical approach to personalized genomics was established, whereby genetic testing performed within a multidisciplinary clinical framework can help inform clinical decision making and guide the selection of appropriate treatment. Follow-up assessment of long-term health
outcomes in relation to implemented treatment could foster research translation and provide evidence in support of clinical utility as an important pre-requisite for the widespread and routine implementation of personalized genetic testing in the clinical domain. Future studies would be facilitated by the easily searchable engine of our research database composed of a growing number of healthy volunteers and patient profiles gathered at a clinical, pathological and genetic testing level. The overall objective is to maximize the South African PSGT platform designed as a clinical tool to identify genomic profiles which can be used to improve risk management.

Despite promising results from on-going research, the greatest determinant of successful implementation of genomic medicine will likely be the willingness of healthcare practitioners to apply genetic knowledge in clinical practice. This necessitates the need for training of a new class of genomic healthcare professionals and development of novel ways to integrate and report the information. A collaborative effort between all stakeholders, including service providers, regulatory bodies, ethics boards, health insurers, clinicians, genetic counselors and patients themselves is thus required via the common thread of pathology. The PSGT approach has helped to inform public healthcare policy regarding the value of microarray-based platforms and is well placed to incorporate next generation sequencing technologies with the potential to improve the standard of care in Africa and elsewhere. Human genome sequencing data linked to clinical information provides a valuable tool for disease prevention, diagnosis and targeted treatment at an individual level.

Ultimately, the use of genetic testing for individual NCD risk assessment and management purposes, irrespective of the number of genes analyzed, will depend on the context within which the test is performed. The value of personalized medicine does not necessarily lie in prediction of risk with 100% accuracy, but rather to provide another line of evidence that can be considered for optimal treatment of patients and to support wellness and health maintenance.130

Acknowledgements

Dr Lynn Horn at Stellenbosch University is acknowledged for valuable advice related to the ethical aspects that had to be taken into account during development of a combined service and research platform. Dr Ruhan Slabbert and Carel van Heerden of the Central Analytical Facility, Stellenbosch University, are thanked for exome sequencing and variant calling using both the hg19 reference sequence and an advanced major allele reference sequence on the same samples. Family practitioner Dr Hein Badenhorst and registered dietitian Lindiwe Whati are thanked for their valuable contribution to the development of the questionnaire and nutrition scores used in the genomics-based chronic disease screening/wellness program. Roberta Rooney is thanked for her valuable contribution to the multiple sclerosis study over many years and Dieter Geiger for research co-ordination. Lujane Nutt, Theresa Brand and Janine Cronje are thanked for assistance with the informed consent process, literature review and administrative support. Dr Mike Urban is acknowledged for helpful discussions and clinical oversight.

Declaration of interest

Research reported in this publication was supported by the Strategic Health Innovation Partnerships (SHIP) Unit of the South African Medical Research Council (MRC) with funds received from the South African Department of Science and Technology (Research grant number 50036665). The MRC is also acknowledged for an internship grant awarded to KE Moremi and a capacity development grant awarded to Y. Yako towards development of the laboratory standard operating procedures in the Pathology Research Facility at Stellenbosch University. This work is also based on the research supported in part by the National Research Foundation (NRF) of South Africa (UID 83962). The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are that of the authors and that the NRF accepts no liability whatsoever in this regard. The Cancer Association of South Africa is acknowledged for a travel grant awarded to doctoral student N van der Merwe. Financial support from Winetech and the Technology for Human Resources and Industry Program (THIRIP) contributed to the development of a Short Course for clinician education registered at Stellenbosch University in October 2014. M. J. Kotze is a director and shareholder of Gknowmix (Pty) Ltd. that has developed a database tool for research translation under the auspices of the Innovation Centre of the South African Medical Research Council (MRC). The Support Program for Industry Innovation is gratefully acknowledged for financial support towards development of the Gknowmix™ Genetic Knowledge Integration System (2007–2009) and Clinical & Scientist Interface Model (2009–2012). In this context, the Technology Innovation Agency is also acknowledged for internship grants awarded to Leslie Fisher, Nicole van der Merwe and Jacobus Pretorius (2013–2014). M. J. Kotze, S. J. van Rensburg, F. J. Cronje and A. V. September are listed as inventors on patents filed by the MRC and the Universities of Stellenbosch and Cape Town. The remaining authors declared no declarations of interest and no writing assistance was obtained in the preparation of this manuscript.

Acknowledgements

Dr Lynn Horn at Stellenbosch University is acknowledged for valuable advice related to the ethical aspects that had to be taken into account during development of a combined service and research platform. Dr Ruhan Slabbert and Carel van Heerden of the Central Analytical Facility, Stellenbosch University, are thanked for exome sequencing and variant calling using both the hg19 reference sequence and an advanced major allele reference sequence on the same samples. Family practitioner Dr Hein Badenhorst and registered dietitian Lindiwe Whati are thanked for their valuable contribution to the development of the questionnaire and nutrition scores used in the genomics-based chronic disease screening/wellness program. Roberta Rooney is thanked for her valuable contribution to the multiple sclerosis study over many years and Dieter Geiger for research co-ordination. Lujane Nutt, Theresa Brand and Janine Cronje are thanked for assistance with the informed consent process, literature review and administrative support. Dr Mike Urban is acknowledged for helpful discussions and clinical oversight.

References


DOI: 10.3109/10408363.2014.997930

Genomic medicine and risk prediction across the disease spectrum

15

Declaration of interest

Research reported in this publication was supported by the Strategic Health Innovation Partnerships (SHIP) Unit of the South African Medical Research Council (MRC) with funds received from the South African Department of Science and Technology (Research grant number 50036665). The MRC is also acknowledged for an internship grant awarded to KE Moremi and a capacity development grant awarded to Y. Yako towards development of the laboratory standard operating procedures in the Pathology Research Facility at Stellenbosch University. This work is also based on the research supported in part by the National Research Foundation (NRF) of South Africa (UID 83962). The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are that of the authors and that the NRF accepts no liability whatsoever in this regard. The Cancer Association of South Africa is acknowledged for a travel grant awarded to doctoral student N van der Merwe. Financial support from Winetech and the Technology for Human Resources and Industry Program (THIRIP) contributed to the development of a Short Course for clinician education registered at Stellenbosch University in October 2014. M. J. Kotze is a director and shareholder of Gknowmix (Pty) Ltd. that has developed a database tool for research translation under the auspices of the Innovation Centre of the South African Medical Research Council (MRC). The Support Program for Industry Innovation is gratefully acknowledged for financial support towards development of the Gknowmix™ Genetic Knowledge Integration System (2007–2009) and Clinical & Scientist Interface Model (2009–2012). In this context, the Technology Innovation Agency is also acknowledged for internship grants awarded to Leslie Fisher, Nicole van der Merwe and Jacobus Pretorius (2013–2014). M. J. Kotze, S. J. van Rensburg, F. J. Cronje and A. V. September are listed as inventors on patents filed by the MRC and the Universities of Stellenbosch and Cape Town. The remaining authors declared no declarations of interest and no writing assistance was obtained in the preparation of this manuscript.

Acknowledgements

Dr Lynn Horn at Stellenbosch University is acknowledged for valuable advice related to the ethical aspects that had to be taken into account during development of a combined service and research platform. Dr Ruhan Slabbert and Carel van Heerden of the Central Analytical Facility, Stellenbosch University, are thanked for exome sequencing and variant calling using both the hg19 reference sequence and an advanced major allele reference sequence on the same samples. Family practitioner Dr Hein Badenhorst and registered dietitian Lindiwe Whati are thanked for their valuable contribution to the development of the questionnaire and nutrition scores used in the genomics-based chronic disease screening/wellness program. Roberta Rooney is thanked for her valuable contribution to the multiple sclerosis study over many years and Dieter Geiger for research co-ordination. Lujane Nutt, Theresa Brand and Janine Cronje are thanked for assistance with the informed consent process, literature review and administrative support. Dr Mike Urban is acknowledged for helpful discussions and clinical oversight.

References


Genomic medicine and risk prediction across the disease spectrum


For personal use only.


