Multigene Next Generation Sequencing Evaluation of the Prevalence of Familial Cancer-Associated Variants in an Arab Population

Musa Alharbi1#, Fawz AlHarthi2#, Nahla Ali Mobark1, Hadeel ElBardis2, Amal Almutairi2, Fatmah Alqubaishi2, Sara AlOtaibi2, Ebtehal Alsolme2, Duna Barakeh2, Wael abdel Rahman Aljabarat1, Abdullah Bany Hamdan4, Ali H. AlFakeeh1, Fatimah Almushawah3, Abdullah A Alsharm4, Ali Abdullah O. Balbaid2 , Naji AlJohani5 ,Alicia Y Zhou5, and Malak Abedalthagafi 2*

1 Department of Pediatric Oncology, Comprehensive Cancer Centre, King Fahad Medical City, Riyadh, Saudi Arabia.
2 Genomics Research Department, Saudi Human Genome Project, King Fahad Medical City and King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia.
3 Department of Surgery, King Fahad Medical City, Riyadh, Kingdom of Saudi Arabia
4 Department of Oncology, King Fahad Medical City, Riyadh, Kingdom of Saudi Arabia.
5 Department of Medicine, King Abdulaziz Medical City, Riyadh, Saudi Arabia.
6 Color Health, Inc. Burlingame, CA, USA

Abstract

In the Kingdom of Saudi Arabia the number of cancer cases increased by 136% from 1999 to 2015, and the average age at presentation of breast cancer in 2013 was 49 years, suggesting a high prevalence of familial breast cancers. An important factor in determining cancer risk, including of breast, colorectal, prostrate, and pancreatic types, is family history. Therefore, a non-invasive diagnostic biomarker screen is needed to predict familial cancer risk in pre-clinical individuals in the Saudi population. This study is one of the first to examine cancer-associated genetic variants within Saudi Arabia using next generation sequencing (NGS). Screening of Arab patients demonstrated a high prevalence of such variations and identified specific variants of importance within this population. Both individuals with and without family history were recruited, at random, from among King Fahad Medical City Familial Genetics Cancer Clinic patients, to be sequenced for mutations using NGS, based on a familial cancer gene panel. A total of 310 subjects, including 123 index patients with cancer and 118 of their family members, 20 of whom also had cancer, were recruited and screened. A set of 536 controls from a database were also screened. Of the 310 subjects, 120 (38.7%) were positive for pathogenic and/or likely pathogenic mutations in *TP53, ATM, CHEK2, CDH1, CDKN2A, BRCA1, BRCA2, PALB2, BRIP1, RAD51D, APC, MLH1, MSH2, MSH6, PMS2, PTEN, NBN/NBS1,* and *MUTYH*. Thus, there is a higher prevalence of genetic mutations for familial cancers in Saudi Arabia than expected based on clinically observed cancer frequencies. Four variants in particular were found to be associated with occurrence of specific cancers in Arabs. One of these was even more prevalent in controls than in subjects, indicating that there are individuals in Saudi Arabia who are at risk for cancer but who are undiagnosed. Hence, we recommend establishing more cancer genetics clinics in Saudi Arabia, training more genetics counsellors in cancer genetics, providing NGS-based testing to identify familial cancer patients at the pre-clinical stage, increasing awareness about familial cancers, and, for patients concerned about passing on a predisposition to familial cancers, offering *in vitro* fertilization and pre-natal genetic diagnosis.

Introduction

The incidence of cancers are on the rise in the Kingdom of Saudi Arabia, increasing by 136% from 1999 to 2015 (Jazieh et al., 2019). An important factor in determining cancer risk and prognosis is family history (Hruban et al., 2010). For example, a 2020 study of colorectal cancer in Saudi Arabia found that colon cancer patients with a family history of colorectal cancer had significantly higher risk of mortality (Azzam et al., 2020). Since the prevalence of familial cancer syndromes and cancer-related mortality in Saudi Arabia appear to be increasing, it is important to uncover which mutations are prevalent in the Saudi population. A better grasp of which variants in which genes contribute to Saudi cancer incidence allows for genetic screening and for recommending preventative measures to those at high risk due to inheritance of mutations. In this study, next generation sequencing (NGS), targeting a selected panel of 30 genes that have been reported to impact the development of breast, ovarian, colorectal, melanoma, pancreatic, prostate, uterine and stomach cancers, was performed on a Saudi population under the umbrella of the Saudi Human Genome Project. We identify genetic variants that might serve as diagnostic biomarkers to predict inherited cancer risk in as yet unaffected or pre-clinical stage cancer patients in the Saudi Arabian population . These results provide important genetic information that can guide future intervention programs to prevent the development or progression of familial cancers. This study contributes to the identification of population-specific genetic cancer risk factors and is expected to have clinical impact on cancer prevention.

Methodology

Subjects and Specimen Handling

The subject population, recruited from the King Fahad Medical City (KFMC) clinics, were patients diagnosed with breast, ovarian, colorectal, melanoma, pancreatic, brain, thyroid, prostate, uterine, and stomach cancers, their family members, and other individuals interested in being screened for mutations believed to be associated with cancer. All these subjects were referred to the Familial Genetic Counseling Clinic. There were no exclusions for age or gender, and, while the majority of participants were Saudi, a small minority were from other Arab countries. A genetic counselor obtained informed consent and pedigree information. Index patients and family members with cancer were placed in a separate cohort for analysis than high risk individuals, defined as first, second, (or, in rare cases, third) degree family members of index patients with cancer. Index patients with neither cancer nor a family history of cancer, along with their family members, were placed in a low-risk category.

Saliva specimens were collected and transported at room temperature using the Oragene DX 510 saliva collection device (DNAgenotek, Kanata, Canada). Samples were forwarded, initially to the Color Genomics Laboratory (Burlingame, CA, USA) by overnight shipping, and later to the King Fahad Medical City Genomic Research Department (Riyadh). Laboratory procedures (Color Genomics) were performed under Clinical Laboratory ImprovementsAmendments compliance (CLIA: #05D2081492).

NGS sequencing

Target enrichment by Agilent's SureSelect method (v1.7) followed by sequencing via Illumina's NextSeq 500 or NovaSeq600 (paired-end 150bp, High Output kit) was done to analyze a panel of 30 genes (*APC, ATM, BAP1, BARD1, BMPR1A, BRCA1, BRCA2, BRIP1, CDH1, CDK4, CDKN2A [p14ARF* and *p16INK4a], CHEK2, EPCAM, GREM1, MITF, MLH1, MSH2, MSH6, MUTYH, NBN, PALB2, PMS2, POLD1, POLE, PTEN, RAD51C, RAD51D, SMAD4, STK11,* and *TP53)* for which mutations have been associated with an elevated risk for breast, ovarian, colorectal, melanoma, pancreatic, prostate, uterine, and stomach cancer.

Methodology

Statistical Analysis

Age distributions of subjects recruited into the study were analyzed by the chi-squared test for association and then by the goodness of fit test for each age range between subject groups. Study participants were classified based on having pathogenic variants, likely pathogenic variants, variants of uncertain significance (VUS), or various combinations of the three , as well categorized into index patients and family members with cancer, high risk relatives, low risk individuals, or controls. A Fisher-Freeman-Halton test for proportional association between subject category and variant class was used to assess disproportionality. Variant types were then collapsed down to the basic three by including individuals with more than one type of variant in the count for each relevant classification, and disproportionally in the number of individuals with the different variant classes was assessed by the chi-squared goodness of fit test. Additionally, the number of pathogenic, likely pathogenic, and VUS variants in each subject or control group was counted and analyzed by the chi-squared goodness of fit test to determine whether one group had a disproportionately high or low number of the different classifications of variants. This latter test was reciprocal to the previous tests, and was conducted to account for the fact that some subjects and controls had more than one variant. Finally, associations between each variant found in subjects and eight types of cancer were analyzed between the index patients and family members with cancer and the cancer-free individuals (i.e. high and low risk subjects and controls), using the Fisher exact test, with p-values adjusted for multiple comparisons (i.e. testing multiple hypotheses of an association of each variant with each cancer). This analysis required an *a priori* comparison of the inheritance of each variant to that of each other variant and of the incidence of each cancer type to each other cancer type to identify dependencies and underlying correlations.

Results

A total of 310 subjects were screened for sequence variants in 30 genes, including 123 index patients with cancer and 118 of their family members, the latter of whom were considered to be high risk individuals, and 20 of whom also had been diagnosed with cancer. The subjects also included 69 low risk individuals (based on lack of cancer family history), including some index patients not diagnosed with cancer. Of the 310 subjects, there were 122 males and 188 females, with an age range of 6 months to 75 years. It should be noted that the age distributions for index patients and family members with cancer, high risk, and low risk individuals were not similar (Figure 1a; p = 0.0025), with the former skewing older (p = 1.16107E-07 for the 40-60 age range and p = 0.0037 for the 60-80 age range). The age distributions were also different for subjects with or without cancer (Figure 1b; p = 0.0038). A set of 536 controls from a database were also screened.

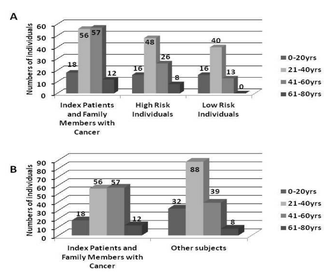

Figure 1: Age distributions in a) different subject groups and b) subjects with and without cancer.

Many of the 310 subjects, as well as some controls, had combinations of multiple pathogenic variants, likely pathogenic variants, or variants of uncertain significance (VUS) (**Figure 2**), including 16 individuals who had more than one pathogenic/likely pathogenic variant. There was a statistically significant association (p = 0.0002) between the variant classifications and the various subject and control groups. In particular, index patients and family members with cancer and high-risk family members without cancer had a higher than expected proportion of pathogenic or likely pathogenic mutations, either alone or in some combination with VUS.

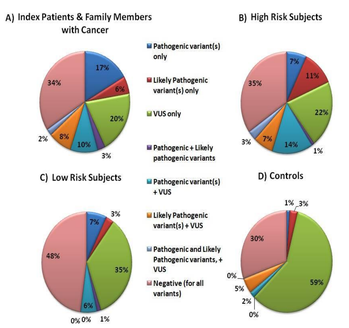

Figure 2: Percentages of subjects and controls with different classifications of variants.

Results

There was a statistical disproportionality in the distribution of patients based on the mutation types they carried (**Table 1**), such that a higher than expected percent of the 123 index patients and 20 family members with cancer had pathogenic (p = 0.000000) or likely pathogenic (p = 0.000006) variants. Likewise, a higher than expected percent of the 69 low risk subjects and the 536 controls, but a lower than expected percent of index patients and family members with cancer and their high risk relatives, had VUS (p = 0.000249). Similarly, the 35 pathogenic variants, 5 likely pathogenic variants, and 82 VUS found in the 310 subjects were distributed in a skewed manner (p = 1.86713E-11), with a disproportionately higher than expected number of pathogenic mutations in index patients and family members with cancer and their high risk relatives (p = 3.16739E-07), and a disproportionately higher than expected number of likely pathogenic mutations in controls (p = 1.54975E-05).

Table 1: Percentages of individuals with pathogenic, likely pathogenic, and/or VUS variants. Individuals with more than one type of variant were included in each relevant variant classification count.

Number of patients with different variant classifications	Index Patients & Family Members With Cancer	High Risk Individuals	Low Risk Individuals	Controls
Pathogenic	31.5%	25.5%	14.5%	3.2%
Likely Pathogenic	18.9%	22.4%	4.4%	7.8%
VUS	40.6%	45.9%	40.6%	65.7%
Pathogenic and/or likely pathogenic, total	45.5%	43.9%	17.4%	10.8%

The co-inheritance of certain variants appeared to be associated, even if the corresponding genes were on different chromosomes and therefore unlinked. This may be because the study participants included families in which one or both parents had multiple variants, with some variants present in both parents, increasing the chances of inheritance of a combination of variants. There was also a highly statistically significant association between the incidences of breast cancer and colorectal cancer (p-value = 0.006, OR = 0.134). Thus, when finding possible associations between the inheritance of each variant and the incidence of each cancer type, the p-values from the statistical analyses had to be adjusted for both the multiple hypotheses being tested and for the variant-to-variant and breast-to-colon cancer associations. Once these adjustments were made, associations were detected between certain variants and specific cancer categories (**Table 2**).

Table 2: Associations of variants with cancer types in the study population:

Variant	Cancer Type	p-value	Odds Ratio
BRCA1 c.5123C>A; p.Ala1708Glu	Breast Cancer	< 0.0001	36.081
BRCA1 c.5123C>A; p.Ala1708Glu	Overall cancer	0.0097	15.35
ATM Δexons 62-63	Prostate Cancer	0.0003	∞[1]
APC c.3920T>A; p.Ile1307Lys	Colorectal Cancers[2]	0.0236	2.996
TP53 c.799C>T; p.Arg267Trp	Brain Cancers[3]	0.025	46.222

1) both of the only two individuals with the variant had prostate cancer, and no cancer-free individuals had the variant in the study population, resulting in an infinite odds ratio
2) including colon, rectal, and sigmoid cancers, as well as polyposis
3) including choroid plexus carcinoma and other CNS tumors

Conclusions

This cohort study investigated genetic variants in a total of 846 patients, their family members, and non-cancer controls in Saudi Arabia. Despite the original variant classifications being based on studies of non-Arab-specific populations, statistically significant associations between some of these variants and cancer occurrence, as well as the disproportionate distribution of pathogenic and likely pathogenic variants to subjects with cancer, indicate that, at least for some pathogenic/likely pathogenic variants, the classifications do apply to those of Arabian ancestry. However, the disproportionately higher number of controls homozygous for VUS, along with a lower than expected percentage of index patients and family members with cancer and their high-risk relatives having a VUS, may indicate that at least some of these variants are benign in those of Arab ethnicity. Nearly half of family members of index patients with cancer (47.5%, i.e. 43 out of 98 high risk relatives and 13 out of 20 with cancer themselves) had at least one pathogenic or likely pathogenic variant, with ten having more than one such variant. Even the control group had an unexpectedly high percentage (10.8%) of individuals with pathogenic or likely pathogenic variants. This greater than expected prevalence of genetic variants with potential cancer risk in the Saudi population suggests a need for increased genetic surveillance programs in Saudi Arabia.

Limitations of this study included a smaller than expected study population, partly resulting from recruitment difficulties due to COVID-19. While this is still the largest study of its type in an Arab population, the relatively small study population did result in extremely small numbers of individuals with certain variants, reducing the statistical analysis power for those variants. A larger study might reveal associations not detected here. For example, it is possible that some variants might be protective, such as APC p.Ser105Gly, APC p.Ser1189Leu, and BRCA2 p.Pro41Leu, which were found more often in controls than in subjects. However, given the very small number of subjects having the APC variants, there was insufficient statistical power to confirm differences in prevalence of these variants relative to cancer status and thereby verify their cancer-protective effects. In contrast, the APC p.Ile1307Lys variant was found to be associated with a higher occurrence of colon cancer, so its disproportionately high prevalence in control individuals may indicate a high frequency of undetected colon cancer in the general Saudi population. Even with the small study population, BRCA1 p.Ala1708Glu showed a statistically significant association with breast cancer (FDR adjusted p-values = 0.0092; OR 35.5). This suggests that Arabic individuals with a family history of breast cancer should be screened for this variant, as it appears to increase the likelihood of developing breast cancer in Arabic individuals. Similar screening should be done for TP53 p.Arg267Trp in families with CNS tumors.

Regarding the observation that the study population skewed towards older individuals in the index patients and family members with cancer, relative to other groups, while this may partly be due to the small sample size, cancer can take years to manifest, even when initiated by inherited mutations. Thus, the skewing may indicate that at least some of the high risk, low risk, or control group individuals with variants also found in the index patients and family members with cancer may develop cancer as they get older, and therefore actually belong in the latter group. Therefore, associations between certain variants (e.g. APC p.Ile1307Lys) and cancer may be underestimated or even undetected in this study, and may be revealed by a longitudinal follow-up study.