GENETIC VARIATION ASSOCIATED WITH UTERINE FIBROID SUSCEPTIBILITY AMONG AFRICAN, ASIAN AND EUROPEAN POPULATIONS.

BY

TUMUHIMBISE PENINAH
Bachelor of Science (biochemistry, chemistry)
15/U13279/PS

A DISSERTATION SUBMITTED TO THE DEPARTMENT OF BIOCHEMISTRY AND SPORTS SCIENCE IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF BACHELOR OF SCIENCE (BIOLOGICAL), MAKERERE UNIVERSITY.

JUNE, 2018
APPROVAL OF DISSERTATION:

This research dissertation has been submitted for review with approval of my supervisor;

Dr. JULIUS MULINDWA PhD
LECTURER,
MAKERERE UNIVERSITY
Department of Biochemistry.

Signature...........................................  Date.................. 5/7/201X
DECLARATION:

I TUMUHIMBISE PENINAH hereby affirm that this dissertation submitted to the Department of Biochemistry, Makerere University has never been presented to any other institute for approval or any award.

Signature                                      Date    July 25 2018
DEDICATION:

I dedicate this work to my mother Ms. Tushemerirwe Joseline for the financial and morale support. May the Almighty God reward you abundantly for all your unwavering endeavors.
ACKNOWLEDGEMENT:

First and foremost, I would like to thank the Almighty God for life and guidance through University.

I would like to express my deep gratitude to my project supervisor, Dr. Julius Mulindwa for all the support and guidance and approval of my research report.

I would also like to thank the staff under Department of Biochemistry for the basic knowledge through lectures that provided me with academic foundation.

Finally, I wish to extend my special thanks to my family and friends for the support and encouragement throughout my course at Makerere University.
ABSTRACT:

Uterine leiomyomata (UL) also known as uterine fibroids are benign neoplasms arising from smooth muscle cells of the myometrium and are clinically recognized in approximately 30% of reproductive age women. (Buttram and Reiter, 1981). Single Nucleotide Polymorphisms (SNPs); chromosome10q24.33 rs7913069, chromosome11p15.5 rs2280543 and chromosome22q13.1 rs12484776 are associated with susceptibility to uterine leiomyoma/fibroids (UL) in Japanese women. The aim of this study was to determine the frequency of those variants in African, Asian and European populations. The study population consisted of 400 individuals from 8 different populations; 50 individuals from African ancestry from South West US (ASW), 50 individuals from Han Chinese in Beijing (CHB), 50 individuals from British in England and Scotland (GBR), 50 individuals from Gambian in Western Division (GWD), 50 individuals from Luhya in Webuye in Kenya (LWK), 50 individuals from Mende in Sierra Leone (MSL) and 50 individuals from Yoruba in Ibadan in Nigeria (YRI). Variant call format (VCF) files from 1000 genomes data Phase3 were used to extract positions and the target populations for further analysis. Asian women were more potentially susceptible to UL because they had the highest Minor Allele Frequency (MAF) in rs2280543 and rs12484776 compared to the other populations. 27% of the CHB population possess rs12484776 and 17% also possess rs2280543. Africans had the second highest MAF for rs2280543 and rs12484776 and Europeans the lowest MAF implying that they are least susceptible to UL. CHB-MSL populations had the highest pairwise weighted $F_{ST}$ estimates which showed that they are unrelated. The Multi-Dimensional Scaling (MDS) plots displayed a large genetic distance between the populations with the lowest being in African populations in which the GWD and LWK were closest together. The large genetic distance indicated high polymorphism and divergence in the populations.
CHAPTER I

INTRODUCTION;

Uterine leiomyomata (UL) also known as uterine fibroids are benign neoplasms arising from smooth muscle cells of the myometrium and are clinically recognized in approximately 30% of reproductive age women. (Buttram and Reiter, 1981).

In eutherian mammals, the uterus supports the development of the embryo and fetus to term and is essential for propagation of the species. The uterus is derived from the embryonic mullerian ducts and is classically divided in to the inner endometrium and the outer myometrium. (Teixera et al, 2008, Spencer et al, 2005). The inner endometrium is composed of luminal epithelium, glandular epithelium and endometrial stroma. The myometrium consists mostly of smooth muscle cells which are organized for maximum efficiency of contraction during labor and supported by an extensive network of large arteries, veins and lymphatic vessels. UL tumors have an estimated incidence of up to 755 in reproductive age women and are usually composed of areas of disordered fascicles of smooth muscle characterized by an excess of acellular extra cellular matrix (ECM). (Gramer and Patel, 1990, Borgfeldt and Andolf, 2000, Baird et al, 2003). Owing to the significant degree of fibroid associated morbidities and to the often ineffective therapies, hysterectomy may be the only option for many women.

Fibroids are clinically detectable in 25%-80% of women in community based studies. Thus many women have the disease without receiving any treatment (Stewart et al, 2013). Majority of the women use herbs, dietary supplements, over the counter medications and alternative medications before seeking out treatment which explains the significant delay in diagnosis of most women. There’s over reliance on hysterectomy in most uterine fibroid patients because they receive educational information from health brochures, friends and family. Uterine fibroids interfere with physical activities of most women for example taking care of their children and going to work, impair relationships with friends and family and gives them a negative sense of their sexuality. Most women are also clearly afraid of future fibroid growth, advancement of the fibroids to cancer and also concerned about the success of their future pregnancies. The prevalence of severe uterine fibroid symptoms ranged from 8%-29% with in African-American women (Stewart et al, 2013)
Twin studies show a strong element of heritability in women undergoing hysterectomy for uterine fibroids. The genetic variants responsible for UL risk can be inherited by offspring. Large linkage studies have the ability to identify common genetic variants of strong effect that contribute to UL risk. Another technique known as admixture mapping can be used to explain how SNPs are associated with susceptibility, admixture mapping is a method of gene mapping that uses a population of mixed ancestry to find the genetic loci that contribute to differences in diseases or other phenotypes found between the different ancestral populations (Shriner et al, 2013).

RESEARCH QUESTIONS;

- What is the frequency of mutations in the SNPs that are associated with susceptibility to uterine leiomyoma among selected African, Asian and European populations?

PROBLEM STATEMENT;

Studies document a 2-3 fold higher incidence of uterine fibroids in African Americans than in European Americans. (Kjerulff, 1993). African Americans tend to have younger ages at diagnosis, more tumors and greater symptomatology than European Americans. (Marshall et al, 1997). There are several clinical and epidemiological observations that suggest that genetic or chromosomal alterations play a significant role in the development of uterine leiomyomata and therefore demonstrate that UL are a clinically and genetically heterogeneous disease.

The current methods of UL treatment for example hysterectomy and drug therapies such as GnRH agonists are limited in their therapeutic effects and therefore remain unsatisfactory. There’s need to identify and choose strategies for fibroid management and optimize the health impact of uterine fibroids.

OBJECTIVES;

GENERAL OBJECTIVE;

- To determine the frequency of SNPs associated with susceptibility to uterine leiomyoma among selected African, Asian and European populations in the 1000 genomes data.

SPECIFIC OBJECTIVES;
➢ To determine the frequency of the mutations in the SNPs associated with UL among selected African, Asian and European populations.
➢ To determine the genetic population structure of the variants in these populations

SIGNIFICANCE;
In this study, the frequency of SNPs already identified as associated with UL was determined (Wise et al, 2012). Using the MAF determined, it is possible to infer potential susceptibility of the different populations to UL. Physicians can be able to customize medical decisions by applying therapies that specifically target genetic alterations in a specific patient population. The effect of drugs with known molecular targets is more efficient and could be of therapeutic advantage in UL treatment.

JUSTIFICATION;
The search/analysis of genetic contributors that cause and also increase uterine leiomyoma susceptibility most especially in African women is necessary. The data from the frequencies of variants can be used in further studies to reduce uterine leiomyoma incidences in African women through healthy nutrition such as consumption of anti-estrogenic foods and medical therapy for example GnRH agonists and SPRMs. The data can also be used to create more awareness about the distribution and susceptibility of uterine fibroids.
CHAPTER II

LITERATURE REVIEW;

Genome wide association study (GWAS) analysis in Caucasian women found the single nucleotide polymorphism (SNP) rs247357 on chromosome 17q253 under one of the suggestive linkage peaks. In East Asian women; three loci on chromosome 10 index SNP rs7913069, chromosome 11 index SNP rs2280543 and chromosome 22 index SNP rs12484776 showed genome wide significant associations.

Uterine leiomyoma;

These benign tumors occur throughout the uterus but are classified by their location in the uterus. Intramural fibroids are located in the myometrial wall. Submucosal fibroids are located at the endometrium-myometrium interface and sub serosal fibroids are located under the serosa of the uterus. Both submucosal and sub serosal fibroids can be pedunculated and extend in to the uterine or abdominal cavity respectively. Intramural, submucosal and sub serosal fibroids can and often exist within the same uterus. The growth rate of multiple tumors in one uterus is not associated with their location or size and concurrent progression and regression of different fibroids may occur under the same hormonal conditions. (Peddada et al, 2008)

Etiology;

Despite the distressing symptoms and prevalence of UL, very little is known about the etiology of these tumors. Uterine leiomyomas are most often diagnosed in perimenopause years, but can become symptomatic much earlier in some women. The incidence of UL declines after menopause. (Walker and Stewart, 2005).

UL incidence increases with age, peaking in the early 40s. However, this could be a result of previously asymptomatic UL becoming more noticeable after years of growth and exposure to endogenous steroid hormones or greater likelihood of older women to seek fertility ending UL treatment20. After menopause, uterine leiomyomas are present in comparable numbers to premenopausal women, however these tumors tend to be smaller and often asymptomatic. Life style choices can have a significant impact on risk of UL development. Obesity, diet, lack of
exercise, and smoking have been correlated with UL incidence. Several studies have shown that high BMI correlates with increased UL incidence. (Ross et al, 1986)

Uterine leiomyomas disproportionately appear in higher percentage in women of African American descent. Even after controlling for BMI, parity, socioeconomic status and other risk factor. African American women have higher incidence, larger tumors at diagnosis, more severe symptoms and earlier age at diagnosis than White, Hispanic or Asian American women. (Marshall et al, 1997). Despite the disproportionate severity and incidence of UL in African American women, the underlying cause for the discrepancy is not well understood.

Karyoptic abnormalities occur in 40-50% of uterine leiomyomas and tumors from the same uterus often show different chromosomal changes. (Bulun, 2013). The most common abnormalities are translocations on chromosome 12, deletion on chromosomes 3q and 7q, trisomy 12 and rearrangement on chromosome 6, 10 and 13. These chromosomal abnormalities may contribute to disruption of genes aberrantly expressed in UL including HGMA2, ESR2 and RAD5. (Walker and Stewart, 2005)

Approximately 25-40% of uterine leiomyomas have simple and nonrandom cytogenetic abnormalities. One of the most common of these abnormalities is a t(12:14)(q15:q23-24) which leaves the coding sequence for the high mobility group (HMG) protein family member (HMGA2) intact but up regulates it’s expression in UL. HMGA2, an non histone component of chromatin and architectural factor, functions to influence transcription and thereby affect diverse cellular processes such as differentiation and proliferation. (Hodge et al, 2008)

**Signs and symptoms;**

- Pain in the abdomen, lower back or pelvis.
- Menstrual; abnormal menstruation, heavy menstruation, irregular menstruation and spotting
- Abdominal distention or cramping
- Frequent urination
- Constipation
- Difficulty emptying bladder
Treatment;

Surgical tools to treat women with uterine fibroids have expanded dramatically in the last decade. Currently, there are four therapies approved by the US Food and Drug Administration (FDA) for treatment of fibroids, Lupron, embolic agents for uterine artery embolization, magnetic resonance imaging-guide focused ultrasound and robotic assisted surgery. Despite the advancement of minimally invasive surgical procedures, hysterectomy remains the mainstay of uterine leiomyoma therapy. (Segars et al, 2014)

Although uterine leiomyomas are a very common gynecologic disease worldwide that can adversely affect women’s reproductive health, fertility and birth outcome, complete understanding of UL pathogenesis remains unclear. Advances in research during the past decade have revealed racial disparities based on molecular differences in leiomyoma development, clinical course and response to treatment. African American ethnicity is considered a risk factor with a higher prevalence than for other races. Further studies are required to develop insight into difference in UL pathogenesis among the different populations. (Catherino et al, 2013)
METHODOLOGY:

MATERIALS:

- Computer - Intel (R) Core™ i5-2520M CPU @ 2.50GHz *64

- Software tools
  - Virtual machine in Oracle Virtual box - Ubuntu 14.04 LTS
  - Vcftools – vcftools_0.1.12b
  - Tabix – tabix-0.2.6
  - Plink – plink-1.07-*86_64
    - Weir and cocker ham’s FST analysis
    - Multi-Dimensional Scaling(MDS)
    - Minor Allele Frequency
  - R and R studio – R *64 3.5.0 and R i386 3.5.0
    - Visualization of MDS clusters.

METHODS:

Study design;

A meta-analysis study was carried out on secondary data extracted from the 1000 Genomes Project Data (https://www.ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/) IGSR: The International Genome Sample Resource. Data analysis was carried out on Variant call files of Chromosomes 10, 11, 17 and 22 as published in phase 3 of 1000 Genomes project.
ANALYSIS OF SNPS, MUTATIONS

Used Ensemble to obtain the positions of SNPs

Obtained data from 1000 genome project
Downloaded 1000 chromosome vcf file

Extracted positions and populations from the vcf files

Removed individuals with missing data
Removed loci with missing SNP data

Determined variant frequencies in the data set

Generated genome file and mds file

R plot of Mds clustering using R studio

Figure 1: Study design
Study samples;
Populations to be analyzed were extracted from 1000 Genomes data. 400 samples were filtered from the European, Asian and African populations (Figure 3). The Asian comprised of 50 Han Chinese in Beijing(CHB) and 50 Punjabi in Lahore(PJL); Europeans comprised of 50 British (GBR) and 50 African ancestry in Southwest US(ASW); African comprised of 50 Gambian in Western Division (GWD), 50 Luhya of Kenya (LWK), 50 Yoruba of Nigeria (YRI) and 50 Mende of Sierra Leone (MSL).

Figure 2: Map from 1000 genomes showing the target populations

Data preparation;
A vcf file for chromosome 22 of phase 3 of 1000 genomes project was downloaded from EBI FTP site (www.ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/) and used to extract the gene loci of the SNPs and also populations that were analyzed.

```
# home/tpeninah/uterine/chr_data/wget
http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/ALL.chr10.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz
```
Vcftools was used to extract populations analyzed from vcf file.

```
Vcftools --gzvcf
ALL.chr10.phase3_shapeit2.mnvcall_integrated_v5a.20130502.genotypes.vcf.gz --keep
POPlist.txt --recode --recode-INFO-all --out chr10_poplist.recode.vcf
```

```
Vcftools --gzvcf
ALL.chr11.phase3_shapeit2.mnvcall_integrated_v5a.20130502.genotypes.vcf.gz --keep
POPlist.txt --recode --recode-INFO-all --out chr11_poplist.recode.vcf
```

```
Vcftools --gzvcf
ALL.chr22.phase3_shapeit2.mnvcall_integrated_v5a.20130502.genotypes.vcf.gz --keep
POPlist.txt --recode --recode-INFO-all --out chr22_poplist.recode.vcf
```
The gene loci of SNPs was extracted using tabix software.

#home/tpeninah/uterine/data/10/ tabix chr10_poplist.recode.vcf.gz 10:105714300-105714300 > chr10_SNPS.vcf

#home/tpeninah/uterine/data/11/ tabix chr11_poplist.recode.vcf.gz 11:203700-203900 > chr11_SNPS.vcf

#home/tpeninah/uterine/data/22/ tabix chr22_poplist.recode.vcf.gz 22:40652800-40653000 > chr22_SNPS.vcf

bgzip –c Chr_SNPS.vcf > chr_SNPS.vcf.gz

Vcftools was used to make plink files (.ped and .map).

#home/tpeninah/uterine/data/10/ vcftools –gzvcf chr10_SNPS.vcf.gz –plink –out chr10_plink

#home/tpeninah/uterine/data/11/ vcftools –gzvcf chr11_SNPS.vcf.gz –plink –out chr11_plink

#home/tpeninah/uterine/data/22/ vcftools –gzvcf chr22_SNPS.vcf.gz –plink –out chr22_plink

Data quality control;

Extracted data was filtered to remove loci with missing data and SNPs with low genotyping rate and those that did not conform to frequency test using plink software. Individuals with missing data were removed. SNPs with low genotyping rate were removed.

#home/tpeninah/uterine/software/plink-1.07-*85_64/ ./plink –noweb --file /home/tpeninah/uterine/data/10/chr10_plink --mind 0.1 --recode --out /home/tpeninah/uterine/data/10/chr10_plink_clean1
Minor allele frequency;

Plink software was used to determine the minor allele frequency of different extracted SNPs that are associated with susceptibility to uterine leiomyomata

MAF for the each population:

Substitute chr10 with chr11 and chr22 for frequency on those chromosomes and change population code accordingly for all the populations.
Population differentiation analysis;
The extent of gene differentiation between populations was assessed using Weir and Cockerham’s F$_{ST}$ analysis method using Vcftools. Pairwise weighted F$_{ST}$ estimate of the SNPs for different populations was determined.

Two by two matrices of pairwise F$_{ST}$ estimate outputs was visualized using MS Excel.

```
#home/tpeninah/uterine/data/10/vcftools --vcf chr10_poplist.recode.vcf --weir-fst-pop ASWPOPList.txt --weir-fst-pop CHBPOPList.txt --out chr10_asw_chb
```

```
#home/tpeninah/uterine/data/11/vcftools --vcf chr11_poplist.recode.vcf --weir-fst-pop ASWPOPList.txt --weir-fst-pop CHBPOPList.txt --out chr11_asw_chb
```

```
#home/tpeninah/uterine/data/22/vcftools --vcf chr22_poplist.recode.vcf --weir-fst-pop ASWPOPList.txt --weir-fst-pop CHBPOPList.txt --out chr22_asw_chb
```

Multidimensional scaling;
The genetic relationships between the populations was determined using multidimensional scaling analysis but a genome file was first generated.

```
#home/tpeninah/uterine/software/plink-1.07-*85_64/.plink --noweb --file
/home/tpeninah/uterine/data/10/chr10_plink_clean2 --genome --out
/home/tpeninah/uterine/data/10/chr10_plink_g
```

```
#home/tpeninah/uterine/software/plink-1.07-*85_64/.plink --noweb --file
/home/tpeninah/uterine/data/11/chr11_plink_clean2 --genome --out
/home/tpeninah/uterine/data/11/chr11_plink_g
```

```
#home/tpeninah/uterine/software/plink-1.07-*85_64/.plink --noweb --file
/home/tpeninah/uterine/data/22/chr22_plink_clean2 --genome --out
/home/tpeninah/uterine/data/22/chr22_plink_g
```
Mds file was then generated.

```
#home/tpeninah/uterine/software/plink-1.07-*85_64/ ./plink --noweb --file
/home/tpeninah/uterine/data/10/chr10_plink_clean2 --read-genome
/home/tpeninah/uterine/data/10/chr10_plink_g.genome --cluster --mds-plot 4 --out
/home/tpeninah/uterine/data/10/chr10_plink_mds

#home/tpeninah/uterine/software/plink-1.07-*85_64/ ./plink --noweb --file
/home/tpeninah/uterine/data/11/chr11_plink_clean2 --read-genome
/home/tpeninah/uterine/data/11/chr11_plink_g.genome --cluster --mds-plot 4 --out
/home/tpeninah/uterine/data/11/chr11_plink_mds

#home/tpeninah/uterine/software/plink-1.07-*85_64/ ./plink --noweb --file
/home/tpeninah/uterine/data/22/chr22_plink_clean2 --read-genome
/home/tpeninah/uterine/data/22/chr22_plink_g.genome --cluster --mds-plot 4 --out
/home/tpeninah/uterine/data/22/chr22_plink_mds
```

Visualization of mds clustering was done using R software.

```
> R

> setwd (“c:/users/vimbah computers/Documents/data”)  

> Q<-read table (“chr10_plink_mds.txt”, header=T)  

> head (Q)  

> plot (Q$C1, Q$C2, pch=20 cex=2, col=”green”)  
```
CHAPTER III

RESULTS:

Minor allele frequency;

The global MAF of rs7913069 is 0.0982/492 which means that 10% of the total population possess the risk variant. In this study, the Mende from Sierra Leone (MSL) have the highest MAF for this variant. MAF of rs7913069 in MSL population (50) is 0.2059/68 which means that 21% of the MSL are associated with susceptibility to uterine leiomyoma.

The global MAF of rs2280543 is 0.0809/405 which means that 8% of the total population possess the risk variant. In this study, the Han Chinese in Beijing (CHB) have the highest MAF for this variant. MAF of rs2280543 in CHB population (50) is 0.1771/96 which means that 18% of the CHB are associated with susceptibility to uterine leiomyoma.

The global MAF of rs12484776 is 0.2326/1165 which means that 23% of the total population possess the risk variant. In this study, the Han Chinese in Beijing (CHB) have the highest MAF for this variant. MAF of rs12484776 in CHB population (50) is 0.2708/96 which means 27% of the CHB are associated with susceptibility to uterine leiomyoma.

The results show that the Han Chinese in Beijing are more susceptible to uterine leiomyoma.

- rs7913069  chr10:105714399;
  Global MAF = 0.0982/492
- rs2280543  chr11:203788;
  Global MAF = 0.0809/405
- rs12484776  chr22:40652873;
  Global MAF = 0.2326/1165
Table 1:

<table>
<thead>
<tr>
<th>POPULATIONS</th>
<th>MINOR ALLELE FREQUENCY</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CHR10 rs7913069</td>
<td>CHR11 rs2280543</td>
</tr>
<tr>
<td>ASW</td>
<td>0.15/60</td>
<td>0.01667/60</td>
</tr>
<tr>
<td>CHB</td>
<td>0.09375/96</td>
<td>0.1771/96</td>
</tr>
<tr>
<td>GBR</td>
<td>0.0125/80</td>
<td>0/80</td>
</tr>
<tr>
<td>GWD</td>
<td>0.1528/72</td>
<td>0.04167/72</td>
</tr>
<tr>
<td>LWK</td>
<td>0.08889/90</td>
<td>0.05556/90</td>
</tr>
<tr>
<td>MSL</td>
<td>0.2059/68</td>
<td>0.05882/68</td>
</tr>
<tr>
<td>PJL</td>
<td>0.1029/68</td>
<td>0.1176/68</td>
</tr>
<tr>
<td>YRI</td>
<td>0.1857/70</td>
<td>0.07143/70</td>
</tr>
</tbody>
</table>

The Minor Allele Frequencies were expressed as percentiles as in the appendix;

![Graphs for CHR10 and CHR11](image)
Population differentiation analysis;

The F<sub>ST</sub> values show population differentiation in the chromosomes of different populations with the highest occurring in CHB-MSL and CHB-YRI. The F<sub>ST</sub> estimates are less than 0.2 which indicates that there’s lack of significant genetic differentiation.

There’s low differentiation between African populations unlike African-Asian and African-European populations.

The 2*2 matrix of pairwise weighted F<sub>ST</sub> estimates of all the populations include;

<table>
<thead>
<tr>
<th></th>
<th>ASW</th>
<th>CHB</th>
<th>GBR</th>
<th>GWD</th>
<th>LWK</th>
<th>MSL</th>
<th>PJL</th>
<th>YRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASW</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>CHB</td>
<td>0.1</td>
<td>0</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>GBR</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>GWD</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>LWK</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>MSL</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>PJL</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>YRI</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ASW</th>
<th>CHB</th>
<th>GBR</th>
<th>GWD</th>
<th>LWK</th>
<th>MSL</th>
<th>PJL</th>
<th>YRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASW</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>CHB</td>
<td>0.1</td>
<td>0</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>GBR</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>GWD</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>LWK</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>MSL</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>PJL</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>YRI</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
</tr>
</tbody>
</table>

**CHR22 rs12484776**

![CHR22 rs12484776](image)
Multi-dimensional scaling;

The MDS is a map that represents the relatedness among populations where the more related are located proximal to one another and the less related further apart.

The Principal Component Analysis plots show that the populations are highly polymorphic and divergent due to the large genetic distance between them.

African populations are more closely related compared to the other populations and have a smaller genetic distance between them.
**DISCUSSION:**
Studies have identified single nucleotide polymorphisms that are associated with susceptibility to uterine leiomyoma and the specific variants (rs7913069, rs2280543 and rs12484776) have been found to have a high MAF in South Asian women. (Wise et al, 2012). The results show that the Han Chinese in Beijing are more susceptible to uterine leiomyoma.

All the $F_{ST}$ estimates are less than 0.2 which indicates that there’s lack of significant genetic differentiation. There’s low genetic differentiation between African populations unlike African-Asian and African-European populations.

The plots show that the populations are highly polymorphic and divergent due to the large genetic distance between them. African populations are more closely related compared to the other populations and have a smaller genetic distance between them.

**CONCLUSION:**
Asian women are more susceptible to uterine leiomyoma then Africans and finally Europeans. This confirms the literature that Asian women are more susceptible to uterine leiomyoma regarding the risk variants investigated. (Wise et al, 2012)

**RECOMMENDATIONS:**
There’s need for further research and more knowledge on uterine leiomyoma in order to minimize the incidence rates and also to prevent the adverse effects of uterine leiomyoma especially hysterectomy.
REFERENCES:


Jenelle C. Hodge, Bradly J. Quade, Mark A. Rubin, Elizabeth A. Stewart, Paola Dal Chin, Cynthia C. Morton, Molecular and Cytogenetic characterization of plexiform leiomyomata
provide further evidence for genetic heterogeneity underlying uterine fibroids. AJP May 2008 Vol 172 No.5.

Kjerulff KH, Guzinski GM, Landenberg PW, Hysterectomy and race, Obstet Gynecol 1993;82(5):757-764


29


Walker CL, Stewart EA. Uterine fibroids: the elephant in the room. Science 2005;308(5728);1589-1592 (PubMed:15947177)

APPENDIX:

VCF File;

VCF file is a text file used in bioinformatics for storing gene sequence variations such as SNPs, insertions, deletions and structural variants. It is usually stored in a compressed manner and can be indexed for fast data retrieval of variants from a range of positions on the reference genome.

<table>
<thead>
<tr>
<th>VARIANT CALL FORMAT FILE</th>
</tr>
</thead>
<tbody>
<tr>
<td>#fileformat=VCFv4.1</td>
</tr>
<tr>
<td>#fileDate=20090805</td>
</tr>
<tr>
<td>#source=myImputationProgramV3.1</td>
</tr>
<tr>
<td>#reference=file:///seq/references/1000GenomesPilot-NCHI36.fasta</td>
</tr>
<tr>
<td>#contig=C/D10,length=62435644,assembly=B36,md5=f126cfdf8a6e0c7437b9d618ff66beb2da,species=&quot;Homo sapiens&quot;,taxonomy=</td>
</tr>
<tr>
<td>#phasing=partial</td>
</tr>
<tr>
<td>#INFO&lt;&lt;ID=NS,Number=1,Type=Integer,Description=&quot;Number of Samples With Data&quot;&gt;</td>
</tr>
<tr>
<td>#INFO&lt;&lt;ID=DP,Number=1,Type=Integer,Description=&quot;Total Depth&quot;&gt;</td>
</tr>
<tr>
<td>#INFO&lt;&lt;ID=AF,Number=1,Type=Float,Description=&quot;Allele Frequency&quot;&gt;</td>
</tr>
<tr>
<td>#INFO&lt;&lt;ID=AA,Number=1,Type=String,Description=&quot;Ancestral Allele&quot;&gt;</td>
</tr>
<tr>
<td>#INFO&lt;&lt;ID=OB,Number=0,Type=Flag,Description=&quot;dSNP membership, build 129&quot;&gt;</td>
</tr>
<tr>
<td>#INFO&lt;&lt;ID=H2,Number=0,Type=Flag,Description=&quot;HapMap2 membership&quot;&gt;</td>
</tr>
<tr>
<td>#FILTER&lt;&lt;ID=q10,Description=&quot;Quality below 10&quot;&gt;</td>
</tr>
<tr>
<td>#FILTER&lt;&lt;ID=e50,Description=&quot;Less than 50% of samples have data&quot;&gt;</td>
</tr>
<tr>
<td>#FORMAT&lt;&lt;ID=GT,Number=1,Type=String,Description=&quot;Genotype&quot;&gt;</td>
</tr>
<tr>
<td>#FORMAT&lt;&lt;ID=GQ,Number=1,Type=Integer,Description=&quot;Genotype Quality&quot;&gt;</td>
</tr>
<tr>
<td>#FORMAT&lt;&lt;ID=DP,Number=1,Type=Integer,Description=&quot;Read Depth&quot;&gt;</td>
</tr>
<tr>
<td>#FORMAT&lt;&lt;ID=H2Q,Number=2,Type=Integer,Description=&quot;HaploType Quality&quot;&gt;</td>
</tr>
<tr>
<td>#CHRCH OS ID REF ALT QUAL FILTER INFO</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>20</td>
</tr>
</tbody>
</table>
Minor Allele Frequency;

The determined minor allele frequency of the different populations was normalized by dividing the MAF by the number of observations to get the MAF RATIO. The MAF RATIO was then multiplied by a hundred to get the percentiles for each population.

<table>
<thead>
<tr>
<th>POPULATIONS</th>
<th>MAF</th>
<th>OBSERVATIONS</th>
<th>MAF RATIO</th>
<th>MAF %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASW</td>
<td>0.15</td>
<td>60</td>
<td>0.0025</td>
<td>0.25</td>
</tr>
<tr>
<td>CHB</td>
<td>0.09375</td>
<td>96</td>
<td>0.000976563</td>
<td>0.09765625</td>
</tr>
<tr>
<td>GBR</td>
<td>0.0125</td>
<td>80</td>
<td>0.00015625</td>
<td>0.015625</td>
</tr>
<tr>
<td>GWD</td>
<td>0.1528</td>
<td>72</td>
<td>0.002122222</td>
<td>0.21222222</td>
</tr>
<tr>
<td>LWK</td>
<td>0.0889</td>
<td>90</td>
<td>0.000987667</td>
<td>0.09876667</td>
</tr>
<tr>
<td>MSL</td>
<td>0.2059</td>
<td>68</td>
<td>0.003027941</td>
<td>0.302794118</td>
</tr>
<tr>
<td>PJL</td>
<td>0.1029</td>
<td>68</td>
<td>0.001513235</td>
<td>0.151323529</td>
</tr>
<tr>
<td>YRI</td>
<td>0.1857</td>
<td>70</td>
<td>0.002652857</td>
<td>0.265285714</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>POPULATIONS</th>
<th>MAF</th>
<th>OBSERVATIONS</th>
<th>MAF RATIO</th>
<th>MAF %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASW</td>
<td>0.01667</td>
<td>60</td>
<td>0.000277833</td>
<td>0.027783333</td>
</tr>
<tr>
<td>CHB</td>
<td>0.1771</td>
<td>96</td>
<td>0.001844792</td>
<td>0.184479167</td>
</tr>
<tr>
<td>GBR</td>
<td>0</td>
<td>80</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GWD</td>
<td>0.04167</td>
<td>72</td>
<td>0.00057875</td>
<td>0.057875</td>
</tr>
<tr>
<td>LWK</td>
<td>0.05556</td>
<td>90</td>
<td>0.000617333</td>
<td>0.061733333</td>
</tr>
<tr>
<td>MSL</td>
<td>0.05882</td>
<td>68</td>
<td>0.000865</td>
<td>0.0865</td>
</tr>
<tr>
<td>PJL</td>
<td>0.1176</td>
<td>68</td>
<td>0.001729412</td>
<td>0.172941176</td>
</tr>
<tr>
<td>YRI</td>
<td>0.07143</td>
<td>70</td>
<td>0.001020429</td>
<td>0.102042857</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>POPULATIONS</th>
<th>MAF</th>
<th>OBSERVATIONS</th>
<th>MAF RATIO</th>
<th>MAF %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASW</td>
<td>0.1167</td>
<td>60</td>
<td>0.001945</td>
<td>0.1945</td>
</tr>
<tr>
<td>CHB</td>
<td>0.2708</td>
<td>96</td>
<td>0.002820833</td>
<td>0.282083333</td>
</tr>
<tr>
<td>GBR</td>
<td>0.2625</td>
<td>80</td>
<td>0.00328125</td>
<td>0.328125</td>
</tr>
<tr>
<td>GWD</td>
<td>0.04167</td>
<td>72</td>
<td>0.00057875</td>
<td>0.057875</td>
</tr>
<tr>
<td>LWK</td>
<td>0.04444</td>
<td>90</td>
<td>0.000493778</td>
<td>0.049377778</td>
</tr>
<tr>
<td>MSL</td>
<td>0.1176</td>
<td>68</td>
<td>0.001729412</td>
<td>0.172941176</td>
</tr>
<tr>
<td>PJL</td>
<td>0.2059</td>
<td>68</td>
<td>0.003027941</td>
<td>0.302794118</td>
</tr>
<tr>
<td>YRI</td>
<td>0.08571</td>
<td>70</td>
<td>0.001224429</td>
<td>0.122442857</td>
</tr>
</tbody>
</table>
Population Differentiation Analysis:

Pairwise weighted $F_{ST}$ estimates;

This represents the 2*2 matrix formed after $F_{ST}$ analysis of the populations. The figures in the matrix represent the pairwise weighted $F_{ST}$ estimates.