Genetic Markers Associated with Anemia in Individuals with Sickle Cell Disease in Tanzania

Liberata Alexander Mwita[1], Raphael Zozimus Sangeda[1], Upendo Masamu[1], David Dynerman[2]

Liberata Alexander Mwita, Raphael Zozimus Sangeda,
Upendo Masamu
Department of Pharmaceutical Microbiology
Muhimbili University of Health and Allied Sciences
Dar-es-salaam, Tanzania
lmwita@blood.ac.tz, sangeda@gmail.com,
umasamu@blood.ac.tz

David Dynerman
The Public Health Company
5662 Calle Real #222
Goleta, California 93117
California, United States of America
emperordali@block-party.net

***Abstract*—Sickle cell disease is a global health problem, a genetic disease which affects many people, particularly common among those whose ancestors came from sub-Saharan Africa. All individuals with sickle cell disease experience anemia which increases the morbidity and mortality. This research aims to identify genetic variants associated with anemia in individuals with sickle cell disease. In the long-term this will contribute towards efforts to improve the life expectancy of individuals by quickly identifying single nucleotide polymorphisms related to anemia in sickle cell disease and enabling better prediction of the severity of anemia that the individual will experience which enable better preventive treatment. Quality control of the Genome Wide Association Studies data and association between anemia and the genotype data were performed using PLINK software and will be presented. Designing of imputation and replication study of the Genome Wide Association Studies data is in progress. The analysis will identify single nucleotide polymorphisms and genes linked to anemia in individuals with sickle cell disease. The results can also be compared with single nucleotide polymorphisms candidates from other studies.**

***Keywords*-Sickle cell disease; Anemia; Genome Wide Association Studies.**

I. INTRODUCTION

Sickle Cell Disease (SCD) is inherited genetic disorder caused by mutation in the hemoglobin (HBB) gene. SCD is a major public health concern [1]. Worldwide, it is estimated that majority of the 275,000 babies born with SCD annually are in sub-Saharan Africa [2]. The burden of SCD in Tanzania is high where it is estimated that 11,000 children are born with SCD annually [3]. SCD causes shortage of healthy Red Blood Cells (RBC) due to the polymerization of the RBCs into a sickle shaped red blood cells. These aggregate in small blood vessels and slow or block blood flow and oxygen initiating vaso-occlusion. Individuals with SCD become anemic because the sickle shaped cells have a short life (10-20 days) unlike normal RBCs which live for 120 days.

Despite the similarity in the origin of the disease, individuals demonstrate varying symptoms and severity. Our previous studies confirmed known and identified new genetic variants associated with fetal hemoglobin [4] and liver function (manuscript write-up ongoing). Other studies [5, 6] have also identified genetic variants associated with different phenotypes observed in individuals with SCD, however much of the variation in phenotype is yet to be explained.

Anemia in SCD increases the morbidity and mortality of individuals. Considering the amount of hemoglobin (Hb) as one variable, non-SCD individuals have a normal range of 13.5-17.5 grams per deciliter (adult men) and 12-15.5 grams per deciliter (adult women). In our database SCD individuals have an average of 8 grams per deciliter. Genome Wide Association Studies (GWAS) involve studying a set of genetic variants in different individuals to see if any variant is associated with a trait by investigating the entire genome of each individual. This study aims to identify genetic variants associated with anemia in individuals with sickle cell disease using a database of GWAS data for 1952 individuals with SCD in Tanzania.

The methodology used to identify the markers will be presented in Section II, followed by the results of the analysis in Section III. Discussion and conclusion of the research will be presented in Sections IV and V, respectively.

II. METHODOLOGY

A. Sampling of subjects and data collection

The phenotype data contains clinical, laboratory and demographic information. Some of these parameters were used in this analysis. Data of 1952 individuals diagnosed with SCD from a cohort have been genotyped. Samples were collected, DNA extracted and genotyped. These individuals are part of the Muhimbili Sickle Cohort recruited at Muhimbili National Hospital, Dar es Salaam, Tanzania. Full details are provided in [4]. Samples were typed on the Illumina Human Omnichip 2.5 platform.

B. Quality control of the genotype data and Association

Standard technical Quality Control (QC) of the data was performed using PLINK software to remove possible sources of technical and genetic bias [7]. This includes removing missing data, duplicates and individuals and Single Nucleotide Polymorphisms (SNPs) failing QC.

 ISBN: 978-1-61208-859-4

Principal Components Analysis (PCA) and the association of the phenotype (Hb) to the QC genotype were done by using PLINK software.

C. Genotype imputation and replication study

Genotype imputation is a statistical inference of unobserved genotypes which is performed on SNPs using known haplotypes in a population such as 1000 Genomes Project in humans. Genotype imputation is underway. Replication in GWAS studies is performed to confirm the phenotype-genotype association results by providing statistical evidence and rule out associations due to biases. Designing of replication and imputation study of GWAS data is in progress.

III. RESULTS

Fig. 1 shows the relationship of the quality-controlled genotype data from our study to other populations.

Our study population (blue dots) is admixture, most of individuals cluster with individuals of African ancestry while few individuals deviate from the cluster. The individuals deviating from the cluster are of Arabic and Indian origin.

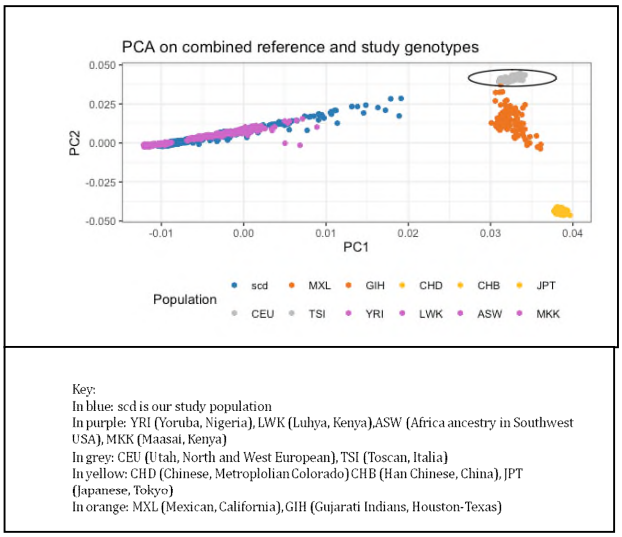

Figure 1. Study population relatedness to other population based on Principal Component Analysis (PCA)

IV. DISCUSSION

The analysis showed that the SNPs associated with anemia are present in the genes that are co-expressed. Individuals with sickle cell anemia experience anemia and frequent infections, this activates the immune response in individuals in order to fight the infections. The SNPs that significantly associated with anemia are found in the genes (Table.1) which function in cell-cell adhesion, antigen receptor-mediated signaling pathway, immune response-activating cell surface receptor signaling pathway and T cell receptor signaling pathway These functions are associated with immune response in humans; it is common for immune system to respond when the human body gets infected.

Other SNPs that significantly associate with anemia are found in the genes (Table.1) that function in hindbrain development and central nervous system neuron differentiation. This is expected in individuals with sickle cell anemia because they experience episodes of pain as well as developmental delays.

Unfortunately, the SNP found to be mostly significant associated with anemia (Fig. 2) at chromosome 3 and 7 have not been annotated hence the functions are not known.

Fig. 2 shows the SNPs (red and blue dots, the p-values on the y-axis) and chromosome in which the SNPs belong on the x-axis.

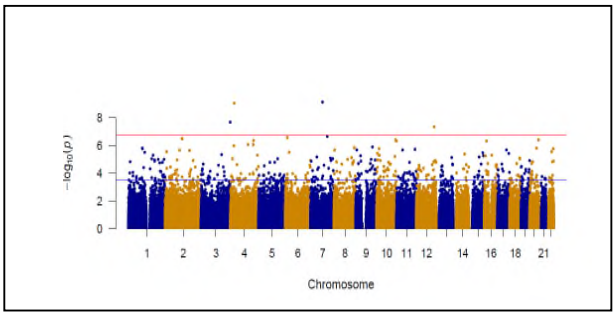

Figure 2. Manhattan plot showing SNPs that associate with anemia in SCD, significant SNPs located at Chromosome 3, 7 and 12.

Some of the SNPs that associate with anemia in individuals with SCD are in Table 1.

TABLE I. FEW SELECTED SNPS AND THEIR LOCATIONS.

SNP	Chromosome	Gene
rs2269688	8	MTMR7
rs11259403	10	PRKCQ
rs13389996	2	CTNNA2
rs10778462	12	CKAP4
rs7136826	12	CLEC1A
rs11632584	15	MEGF11
rs7163369	15	SLCO3A1
rs732523	12	PCED1B
rs17276467	7	CREB3L2
rs10209276	2	KCNH7
rs4578863	2	ZC3H6

 ISBN: 978-1-61208-859-4

It is our hope that the completion of replication and imputation analysis will reveal more and significant associations.

V. CONCLUSION AND FUTURE WORK

This study indicated genetic markers (SNPs) that associate with anemia in individuals with SCD. This is the first step towards developing a tool that will quickly identify the markers linked to anemia in SCD individuals which is an important step in improving preventive treatment of these individuals. Similar analysis has to be extended in same and different sickle cell disease cohorts in order to identify new and confirm the variants linked to anemia in individuals with SCD.

ACKNOWLEDGMENT

This research is sponsored by University of California Global Health Institute (UCGHI) GloCal Health Fellowship.

REFERENCES

[1] Un.org, “Secretary-General's message on sickle-cell anaemia | United Nations Secretary-General” [Online] Available from: https://www.un.org/sg/en/content/sg/statement/2009-0619/secretary-generals-message-sickle-cell-anaemia, 2009, [Accessed 15 Nov. 2018].

[2] E. Ambrose et al., “High birth prevalence of sickle cell disease in Northwestern Tanzania,” Pediatric Blood & Cancer, vol. 65(1), pbc.26735, Jan. 2018, doi: 10.1002/pbc.26735.

[3] F. Tluway and J. Makani, “Sickle cell disease in Africa: an overview of the integrated approach to health, research, education and advocacy in Tanzania, 2004-2016, “British Journal of Haematology,” vol. 177(6), pp.919-929, June. 2017, doi: 10.1111/bjh.14594.

[4] S. Mtatiro et al., “Genome wide association study of fetal hemoglobin in sickle cell anemia in Tanzania,” PLoS ONE, vol. 9(11), pp.e111464, Nov. 2014, doi: https://doi.org/10.1371/journal.pone.0111464

[5] P. Sebastiani et al., “Genetic modifiers of the severity of sickle cell anemia identified through a genome-wide association study,” American Journal of Hematology, vol. 85(1), pp.29-35, Jan. 2010, doi: 10.1002/ajh.21572.

[6] L. Liu et al., “Original Research: A case-control genome-wide association study identifies genetic modifiers of fetal hemoglobin in sickle cell disease,” Experimental Biology and Medicine, vol. 241(7), pp.706-718, April, 2016, doi: 10.1177/1535370216642047

[7] S. Purcell et al., “PLINK: a tool set for whole-genome association and population-based linkage analyses,” The American Journal of Human Genetics, vol. 81(3), pp.559-575, Sep. 2007, doi: 10.1086/519795.

 ISBN: 978-1-61208-859-4