WHOLE EXOME SEQUENCING TO IDENTIFY MUTATIONS IN GENES IN NEPALESE PATIENTS WITH RARE BLEEDING DISORDERS

M.Sc. Thesis

2017

Submitted to

Central Department of Biotechnology

Tribhuvan University

Kirtipur, Kathmandu, Nepal

By

Binod Neupane

Registration No: 5-2-37-963-2009

Supervisors

Dr. Tilak R. Shrestha
Professor
Central Department of Biotechnology,
Tribhuvan University, Kirtipur,
Nepal

Dr. Sridhar Sivasubbu
Senior Scientist
CSIR-Institute of Genomics and
Integrative Biology, New Delhi,
India

Acknowledgement

Though only my name appears on the cover of this dissertation, a lot of guidance and assistance from many great people have contributed to its success and final outcome. I owe my gratitude to all those people who have made this dissertation possible and because of whom my dissertation experience has been one that I will cherish forever. Whatever I have done is only due to such guidance and assistance and it's my great pleasure to thank them.

First and foremost, I would like to express my profound gratitude to my thesis supervisor Prof. Dr. Tilak R. Shrestha for his eternal support, proper guidance and constant feedback without which this research would not have been done. The present research work on rare bleeding disorders is based on his ideas, preconception and collaborative research agreement done with Dr. Shridhar of IGIB, New Delhi, India.

I am equally indebted to my another supervisor, Dr. Sridhar Sivasubbu for accepting me as a training fellow and providing me an opportunity to access the world class laboratory and research facilities including Next Generation Sequencing and Zebra fish facilities. His guidance, encouragement and dedication has not only helped me at the time of this research but also has helped me in personality development.

My deep appreciation goes out to Prof. Dr. Rajani Malla, former HOD of CDBT-TU (HOD at our time) for her kind support my thesis work.

I am extremely thankful to NGS teammates of SSB lab (Lab No. 123) of CSIRI-IGIB viz. Shamsudheen K Vellarikkal, Ankit Verma, Rijith Jayarayan, Rowmika Ravi and Anoop Kumar for sharing expertise, sincere and valuable guidance and encouragement extended to me.

My deep appreciation goes out to members of Nepal Hemophilia Society (NHS) especially President Mr. Bed Raj Dhungana, Secretory Mr. Ujjawal K.C., Program manager Ms. Laxmi Karki and nurses working at Hemophilia Care Unit, Bir Hospital. Similarly, all the patients with rare bleeding disorders and their family members who participated in this research without whom my dream to work in molecular characterization of Nepalese Hemopiliacs would have never been completed, owe my deepest acknowledgement.

I owe my special thanks to all my batch mates (CDBT 5th Batch); professors and lecturers at CDBT-TU; my seniors and my juniors. Medha K.C., Nutan Thakur, Sujan Biswakarma and Gauri Thapa deserve more gratitude for their kind support throughout my whole M.Sc. including this research work.

Last but not the least, I feel very happy to thank my ever inspiring mom-dad, sisters and brothers for their eternal love, care and encouragement. They have been supporting me all through the thick and thin from my birth and always have kept me in high spirit throughout the entire period of my academic as well as personal life.

Abstract

Rare bleeding disorders (RBDs) are among the oldest described genetic diseases, generally leading to lifelong hemorrhagic complications. These are monogenic in nature and are inherited in Mendelian patterns. The genetic cause of RBDs is the defect(s) in gene(s) coding or regulating various clotting factor(s). RBDs manifest themselves in the form of either severe or moderately severe or mild and have affected approximately 400,000 individuals worldwide. Von Willebrand disease and hemophilia A are the most common type of RBDs. Since the clinical presentations of various types of RBDs intersect with each other, only the laboratory studies may not be sufficient for the accurate diagnosis of the RBDs. Genetic studies are required in such cases. Moreover, genetic studies allow better understanding of the biology of rare bleeding disorders and the genetic information can be used for the translational application, prenatal diagnosis and the detection of carrier status, prediction of development of inhibitors and can also assist in genetic counseling. However, traditional molecular techniques have shown limitations in efficient characterization of mutations causing RBDs. In present era of high through-put sequencing, Next Generation Sequencing (whole genome sequencing and whole exome sequencing) which has emerged as a gold-standard for the identification of disease-causing mutations in various other rare Mendelian diseases has also shown a convincing potential to explore the underlying genetic lesions in the patients with rare bleeding disorders. In our current study whole exome sequencing has been used for the screening of mutations in patients suffering from two rare bleeding disorders viz. Type 2 Normandy von Willebrand disease (2N VWD) and Factor X Deficiency (FXD). Sequencing was performed in Illumina platform (HiSeq 2500). We developed our own bioinformatics analysis pipeline for WES data and ended up with only one causative mutation in both the RBDs following rigorous prioritization of the variants. The causative mutation identified in FXD, c.T212C:p.F71S, which is reported as a founder effect in Algerian population has not yet been reported from the other parts of the world. In case of 2N VWD, the causative mutation identified, c.C2446T:p.R816W is one of a very common variant reported all over the world. Both the causative mutations were validated by capillary sequencing and also the carrier status among the family members was checked. We found two daughters of male patient of 2N VWD are carrier for the disorder.

Key words: rare bleeding disorders, 2N VWD, FXD, whole exome sequencing, bioinformatics analysis of WES data, validation of WES results, detection of carrier status

TABLE OF CONTENTS

List of Figures

List of Tables

List of Abbreviations

%	-	Percentage
2N VWD	-	Type 2 Normandy von Willebrand disease
AD	-	Autosomal Dominant
ADAMTS13	-	A Disintegrin And Metalloproteinase with a Thrombospondin Type 1 Motif, Member 13
APTT	-	Activated Partial Thromboplastin Time
AR	-	Autosomal Recessive
B	-	Benign
BAM	-	Binary Alignment/Map
bcl	-	base call
BT	-	Bleeding Time
BWA	-	Burrow Wheeler Aligner
BWA-MEM	-	Burrows-Wheeler Alignment- Maximal Exact Matches
CASAVA	-	Consensus Assessment of Sequence And Variation
Chr	-	Chromosome
CSGE	-	Conformation Sensitive Gel Electrophoresis
D	-	Deleterious or Probably Damaging
ddNTPs	-	DiDeoxy Nucleotides
DDW	-	Double Distilled Water
DGGE	-	Denaturing Gradient Gel Electrophoresis
DHPLC	-	Denaturing High Pressure Liquid Chromatography
EDTA	-	Ethylene Diamine Tetra-Acetic Acid
ERGIC-53	-	Endoplasmic Reticulum-Golgi Intermediate Compartment 53 Kda Protein
EtOH	-	Ethyl Alcohol
ExAC	-	Exome Aggregation Consortium
F	-	Phenylalanine
F10	-	Coagulation Factor X gene
F11	-	Coagulation Factor XI gene
F12	-	Coagulation Factor XII gene
F13A1	-	Coagulation Factor XIII A Chain gene

F13B	-	Coagulation Factor XIII B Chain gene
F2	-	Coagulation Factor II gene
F5	-	Coagulation Factor V gene
F7	-	Coagulation Factor VII gene
F8	-	Coagulation Factor VIII gene
F9	-	Coagulation Factor IX gene
FC	-	Flow Cell
FGA	-	Fibrinogen Alpha Chain
FGB	-	Fibrinogen Beta Chain
Fig	-	Figure
FII	-	Factor II
FIX	-	Factor IX
FV	-	Factor V
FVIII	-	Factor VIII
FX	-	Factor X
FXD	-	Factor X Deficiency
FXI	-	Factor XI
FXII	-	Factor XII
gDNA	-	Genomic Deoxy Ribonucleic Acid
GGCX	-	γ- Glutamyl Carboxylase gene
GWAS	-	Genome Wide Association Studies
HA	-	Hemophilia A
HB	-	Hemophilia B
HMWK	-	High Molecular Weight Kininogen
IGV	-	Integrative Genomics Viewer
Indels	-	insertions deletions
IU/ml	-	International Units per Millilitre
LMAN1	-	Lectin Mannose Binding 1
mg	-	Miligram
mL	-	Mililitre
MCFD2	-	Multiple Coagulation Factor Deficiency 2

mRNA	-	Messenger Ribonucleic Acid
NCBI	-	National Center for Biotechnology Information
NFW	-	Nuclease Free Water
NGS	-	Next Generation Sequencing
NHS	-	Nepal Hemophilia Society
PCR	-	Polymerase Chain Reaction
Polyphen2	-	Polymorphism Phenotyping v2
PT	-	Prothrombin Time
Q	-	Phred Quality Score
R	-	Arginine amino acid
RBDs	-	Rare Bleeding Disorders
RSB	-	Resuspension Buffer
RT	-	Room Temperature
S	-	Serine
SAM	-	Sequence Alignment/Map
SDS	-	Sodium Dodecyl Sulfate
SIFT	-	Sorting Intolerant From Tolerant
SNVs	-	Single Nucleotide Variations
SSCP	-	Single Strand Conformation Polymorphism
TAE	-	Tris-Acetate-EDTA
TE	-	Tris-EDTA
UCSC	-	University of California, Santa Cruz
UTR	-	Untranslated Region
vcf	-	variant call format
VKORC1	-	Vitamin K Epoxide Reductase 1 gene
VWD	-	von Willebrand Disease
VWF	-	von Willebrand Factor
W	-	Tryptophan amino acid
WES	-	Whole Exome Sequencing
WGS	-	Whole Genome Sequencing
WFH	-	World Federation of Hemophilia