Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2021 doi:10.20944/preprints202103.0088.v1

A family based whole exome sequence study to indentify modifier genes for phenotype heterogeneity between severe and non-severe thalassemia patients

*Dipankar Saha[a], Prosanto Kumar Chowdhury [a], Debashis Pal[a], Kaustav Nayek[d], Gispati Chakraborty[e], Surupa Basu[b], Anupam Basu[a] **

[a] Department of Zoology, The University of Burdwan, PurboBarddhaman- 713104, West Bengal, India. [b] Institute of Child Health, Kolkata, [c] Thalassemia Control Unit, Burdwan Medical College and Hospital, [d]epartment of Paediatric Medicine, Burdwan Medical College and Hospital, [e] Burdwan University Health Centre, The University of Burdwan,

Abstract

Whole Expanded Exome Sequencing Study of two families father, mother and index cases (trio) was undertaken for two E/ beta thalassemia subjects with same HBB genotype. Approximately 200ng of DNA was taken from each individual and shared into 300-400bp fragment. Then the shared fragments are end repair. Klenow exonuclease was used to add an adapter. After adapter ligation 10 cycle PCR amplification was done for each sample. The targeted Exome was captured by the Agilent Sure select XT Human all Exome V6+ UTR kit as per the manufacturer's protocol. Captured library was then amplified 10 cycles with 8 bp index sequence for each sample. Then the indexed capture library was pooled together. Pair end sequencing of the pooled library was performed in Illumina HiSeq2500 using Illumina HiSeq SBS kit. Finally, both the genes, inherited and denovo, from both the subjects were separately functionally annotated by DAVID online tools 6.8. Functionally annotation result shows that in case of subject-1, 6 KEGG pathway were involved. These are Adherent junction, Protein digestion and absorption, Inflamatory Bowel Disease, Amoebiasis, PPAR signaling pathway and glycolysis or gluconeogenesis. Interestingly in case of subject-2, only 2 KEGG pathway were found, Thyroid hormone synthesis and carbon metabolism.

Keywords: WES; Thalassemia; phenotype

Introduction:

Thalassemia is a monogenic disease caused by the mutation in the beta globin (HBB) gene. There are more than 400 disease causing mutation is responsible for the causing beta thalassemia worldwide. (hbvar ref). HbE beta thalassemia is a compound heterozygous mutation of the HBB gene, found in the South Esat Asia. In Eastern India there are almost 17 HBB mutation are responsible for the disease. The most common mutations are IVS 1-5(G>C) [HBB:c.92+5G>C] and CD26(G>A) [HBB:c.79G>A], responsible for the HbE- beta thalassemia (1). The clinical severity of the HbE-beta thalassemia is varying from very severe to very mild. Some subjects are presenting the disease at very early age and needing regular transfusion for survival, but some are presenting late onset and not requiring regular transfusion or required occasional transfusion (2, 3).Several studies are done to find out the factors responsible for the clinical severity of the HbE-beta thalassemia subjects. Several studies are showing the modifiers like beta mutation type, fetal hemoglobin level, genetic polymorphism, like -158(C>T) on Gγ globin gene(4), HBS1L-MYB intergenic polymorphism, BCL11A gene polymorphism,(5,6,7) responsible for the HbF inducer and also the alpha globin gene deletions. The result shows some are

* Corresponding author: Prof. Anupam Basu, Department of Zoology, The University of Burdwan, PurboBarddhaman, West Bengal- 713104, India, Email- abasu@zoo.buruniv.ac.in

a *&* ***b*** *are equal contribution*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2021 doi:10.20944/preprints202103.0088.v1

statistically significant, but this is not enough to understand the reason of such clinical diversity. So, we hypothesized that there are several other genetic loci, other than the globin cluster are present, which may modify the disease severity in the different way other than the modulation of the globin production.

To finding out the genetic modifiers, responsible for the disease severity of the HbE-Beta thalassemia subject, with similar genotype [IVS1-5(G>C)/CD16(G>A)] we did Whole Expanded Exom Sequencing Study of two family (father, mother and affected patients).

Materials and Methods:

Subjects Information: We select two HbE-Beta thalassemia subjects with same HBB genotype IVS1-5(G>C) and CD26(G>A) with different phenotype: One is transfusion dependent (TD) and another one is non transfusion dependent (NTD). To check whether any de novo mutations have any role as modifier loci, parents of the both the subjects were also included in this study.. Clinicopathological information of both the subjects have been presented table-1.

Subject-1, which is TDT present the disease at 10 month of age and taking regular transfusion of 2-month interval for last 13 year. Steady state haemoglobin level was 6.1 g/dl . Subject- 2, which is NTDT present the disease at 14 years of age. She has not taking any transfusion for last 16 years. Steady state haemoglobin was 7.8 g/dl. She has a splenomegaly of 9 cm from the left costal margin.

A total of 6 subjects were included in this study as described above including both the parents of each index cases. The study has been approved by the Institutional Cinical Ethics Committee, The university of Burdwan

Sample Collection and DNA extraction: Pre transfused 3 ml peripheral blood samples were collected from individual subjects in EDTA from each subjects with proper phlebotomy procedure. DNA was extracted by using the commercial DNA extraction kit.

Whole Exome sequencing:

Approximately 200ng of DNA was taken from each individual and shared into 300-400bp fragment. Then the shared fragments are end repair. Klenow exonuclease was used to add an adapter. After adapter ligation 10 cycle PCR amplification was done for each sample. The targeted Exome was captured by the Agilent Sure select XT Human all Exome V6+ UTR kit as per the manufacturer's protocol. Captured library was then amplified 10 cycles with 8 bp index sequence for each sample. Then the indexed capture library was pooled together. Pair end sequencing of the pooled library was performed in Illumina HiSeq2500 using Illumina HiSeq SBS kit V4 as per manufacturer's protocol. Generated FASTQ file was used for the variant calling and in house data analysis.

WES Data Analysis:

Variant calling: All the raw reads were checked and aligned with hg19 human reference genome using Burrows- Wheeler Aligner (BWA) and variant call was made by Genomic Analysis Tool Kit (GATK). The VCF was annotated by using the VarAFT tools.

Hunting of responsible variants and gene for clinical significance:

De novo variants: Variants of the Index cases were compared with the variants of the father and mother and inherited variants were filtered out.

Inherited variants: After annotating the VCF file using VarAFT tools we filter out all Upstream, downstream, intergenic, UTR, nocoding and non splicing related intronic variants, only exonic and splicing variants are considered for further analysis. To hunt the responsible Inherited variants two broad scheme were applied. In the first scheme, only minor variants were targeted. In the second scheme, homozygous pathogenic variants were targeted as follows:

Scheme 1: We considered only non-synonymous exonic variant and splicing variants with frequency of <0.01 (1%) in 1000 Genome database, for the next step of analysis. We align the individual's variants with the father, mother variants to detect the inherited and *de novo* variants. The variants were again filter based on the SIFT and PolyPhen2 score. The variants showing

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2021 doi:10.20944/preprints202103.0088.v1

pathogenic either any of one were taken for the final variants. Then the variants were maped into gene and the genes were functionally annotated with the DAVID online tools 6.8 to searching out the biological pathway, KEGG pathway involved with these genes.

Scheme 2: All the synonymous, unknown and heterozygous variants are further filtrated. To predict the effect of mutation, SIFT and POLYPHEN2 scoring were applied on all the homozygous exonic and splice variants only deleterious variants by this two software were considered. The genes containing these variants are further functionally clustered by KEGG path way using DAVID online tools 6.8.

Results

A total of 91194 variants were identified in the **TDT subject** and 91196, 90530 variants were identified in the father and mother respectively. On the other hand a total of 90226 variants were identified in the N**TDT subject** and 89396, 88275 variants were identified in the father and mother respectively.

De novo Variants

After filtering out the finally we got the 520, 551 and 503 non synonymous exonic and splicing variants in subject-1, father and mother respectively. Out of 520 variants, 514 are inherited variants and only 6 are denovo variants for **TDT subject** . On the other hand, 538,530 and 541 variants were found in **NTDT subject**, father and mother respectively. Out of 538 total variants in index case, 534 were inherited variants and 4 were de novo variants (Table 2) .

Inherited variants and functional annotation:

As mentioned in the method section, for hunting of the functional clinical responsible variants and genes, using minor allele approach (Fig 1) and homozygous variants approach (Fig 2) available genes and variants are listed in Table 3 -6.

Discussion

We performed a trio based study with 2 beta haemoglobinopathy patients of same age, sex, and harbouring the same primary mutation, however, one was transfusion-dependent (TDT) and the other non-transfusion-dependent (NTDT). We focussed on single-nucleotide variations and short indels in a whole-exome sequencing approach, and then tried to explain their differing clinical phenotypes through their differing mutations with expected most severe impact. After associating mutations of gene products to the metabolic pathways they have most significant roles in, we found that the mutations and their related individual phenotypes belonged to either of 2 categories – those related to general defects of growth and development, and those related to anaemia, either by reduced/ineffective erythropoiesis or by increased RBC haemolysis.

In our cases of interest, we found that mutations of genes associated with growth and development were present in both patients, while those possibly involved in erythrocyte membrane or even general cytoskeletal integrity were affected in only the TDT individual. Also interesting was that the NTDT patient displayed a much lower mutational load than the TDT patient (about 1/3rd of the TDT patient's mutational load).

Patient 1 (addressed as the TDT patient), was diagnosed at 10 months of age and was on once in 2-3 months transfusion regimen from then. She was maintaining and average pre transfusion haemoglobin level of 6.05 gm/dl. Her height and weight for age were both below the 3rd centile against age and sex matched subjects. Both the liver and spleen were palpably enlarged to the extent of 4.5 and 7 cm below the right and left costal margins respectively. The level of lactate Dehydrogenase, serum haptoglobin and unconjugated bilirubin remained significantly above normal range, during this period (8). Provisionally attributing the increase of such parameters to excessive ineffective erythropoiesis she was put on regular monthly transfusion from 7 years of age. Attempting to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2021 doi:10.20944/preprints202103.0088.v1

maintain average pre-transfusion haemoglobin level of above 9 gm/dl with 15 ml/kg of 60% packed cell transfusion. Though the biochemical parameters of ineffective erythropoiesis improved but never touched baseline, this hinting at intravascular haemolysis as a probable cause (9). Upon starting on the new transfusion regime, minimal improvement in growth was noted. Though the average pretransfusion haemoglobin level remained around 8 gm/dl, with 3 weeks transfusion interval regime. The iron overload increased at a rate more than expected. After follow up till 18 years of age, it was noted that she was having a number of comorbidities, not unusual for thalassaemics but at an earlier age. She was detected as bio-chemically hypothyroid with free tri-iodothyronine level of 0.8 ng/ml and Thyroid stimulating hormone level of 7.02 ng/ml, which was corrected with levothyroxine replacement therapy. She had thelarche at 16 years and menarche at 18 years of age, both delayed by about 2 years, following the secular trend. The bone age at 18 years was lagging by 2 years and the Z Score of bone mineral density at trochanteric level was -1.0, signifying Osteopenia. She was detected as having multiple small gall bladder stones at the same age. We also contributed this delay of catch up growth and development to a chronic allergic state and chronic ill health due frequent abdominal cramps and diarrhoea leading to malabsorption and micronutrient deficiency. After replacement therapy with essential mineral (calcium and zinc) and vitamin (Vit D) replacement and probiotics, her condition has improved, but her growth still remains below the 3rd centile with moderately enlarged liver and spleen.

After the data of the NGS was made available, it was noted that she was compound heterozygous for an ACTN3 mutation, which is thought to be involved in attachment of actin filaments to intracellular structures. This mutation could have far-reaching consequences, though in RBC it would most likely lead to structural destabilization and increased erythrolysis, thus exacerbating the anaemic condition and enforcing a transfusion-dependent state (10). Also present was a mutation in FARP2, a guanine exchange factor involved in cytoskeletal remodelling by RAC1, which has been implicated in erythrocyte maturation (12) and is known to be dysregulated in sickling RBCs (11). These mutations could explain the extra burden of ineffective erythropoiesis and intravascular haemolysis, evidenced by hyperbilirubinemia, high LDH and Haptoglobin and compensatory increase in size of liver and spleen and formation of gall stones at such an early age. Outside of anaemic conditions, though, there were mutations identified in both patients relating to various pathways of growth and development, with a common motif being genes involved in cellular respiration and fatty acid metabolism. Outside of genes involved in carbon metabolism, the TDT subject possessed mutations in IL12RB1 and TLR4, which may be contributory to defective T Helper cell-mediated immune response and gamma interferon production and its related antiviral response (15) – the propensity of this subject to fall sick very often, mostly to influenza like viral illness, wrongly interpreted as allergic hyperactive airway disorder (14, 15). More over IL12RB1 mutation may be responsible for inflammatory bowel disease like state (13, 16), which could be responsible for the malabsorption like symptoms leading to lactose intolerance and micronutrient deficiency. Reduced bone mineral density and resultant osteopenia may, among other things, be attributable to the menagerie of collagen deficiencies. (17).

Learning that this patient also presented with very early inset hypothyroidism, which is unlikely in well chelated thalassaemic individuals association with homozygous mutations in the HLA-DR3, HLA-DPB1 and the TPO genes may be corroborative. It has been shown that individuals with an amino acid substitution at position 74 of the DR beta 1 chain of HLA-DR3 (DRb1-Arg74), susceptibility to autoimmune thyroid disorder increases [18]. The analysis of amino acid variants of HLA molecules of HLA-DPB1 were strongly associated with Graves Disease, especially amino-acid signatures of the HLA-DP β chain, might contribute to the molecular pathogenesis of early-onset auto-immune thyroid disease (AITD) [19]. *TPO* gene mutations result in disruption of thyroid hormone synthesis and are classified as thyroid dyshormonogenesis. The combined effects of these genetic aberrations may have contributed to the individual presenting as clinically and biochemically hypothyroid at such an early age, to require thyroid hormone replacement therapy. [20] Presence of homozygous mutation in the FUT3 gene may be of significance in relation to increased

transfusion frequency. Mutation in FUT3 gene is actually reflected as the gene for Lewis Antigen (Le^{a+} or Le^{b+}) a minor blood group antigen, so the person who will not be able to synthesize the antigen (protein product of the Le gene) and will be Le Negative (Le^{a-} Le^{b-}), as the transfusable blood comprising of packed red cells are generally not screened for such minor blood groups, to in individuals who receive frequent transfusions, may develop Anti Lewis antibodies to such seemingly innocuous antigen, thus causing partial haemolysis of the transfused blood which is positive for such antigen. This small amount of haemolysis is not severe enough to cause catastrophic transfusion reaction, but will decrease the transfusion interval subtly, increasing the levels of bilirubin and LDH. This could be another reason for increased haemolysis which persisted even after starting the regular transfusions, which should have ideally got rid of the endogenously produced defective red cells. Generally, in sporadically transfused individuals Lewis blood group is seldom responsible for such haemolysis, however in multi and regular transfusion scenarios, it compromises transfusion efficiency, hence decreasing the frequency between transfusions. [21] Though the mutations in the FUT3 may be associated with H. Pylori infection causing frequent abdominal cramps, etc, this has not been substantiated in our subject. [Ref Table 4]

Apart from a considerably larger mutational load on fatty acid metabolic pathway genes in the TDT patient, thyroid hormone production was almost equally affected in both. Other than this pathway in the NTDT patient, angiogenesis, limb and connective tissue (collagen) formation and immune development were severely affected in the TDT individual, sometimes even multifactorial.

When, we compared the clinical condition of Index case 2, the NTDT, it was certainly different. She presented at the clinical set up at the age of 12 years, with and average baseline haemoglobin of 6.0 gm/dl, with no hepatomegaly and palpable splenomegaly of 6 cm below the left costal margin. The LDH, unconjugated bilirubin and haptoglobin were only more than 2×upper limit of normal (ULN), rest of the parameters being normal and documenting only 20% Nucleated red cells in peripheral circulation. These findings are truly correlated with the absence of mutations contributing defects in the RBC cytoskeleton, when compared to the TDT subject.

She was detected as suffering from biochemical hypothyroidism at the age of 11 years with TSH of 6.0 ng.ml and was put on thyroid replacement therapy had spontaneous thelarche at 12 years and menarche at the age of 12 years 7 months. She has not been put on regular transfusion regimen, though her height for age was at 3rd centile, as her mid-parental height was also at the third centile and her growth velocity was satisfactory. There was definite increase in average baseline haemoglobin and growth spurt after thyroid hormone replacement therapy was initiated. The NGS data verifies the rampant hypothyroidism by noting mutations in thyroglobulin and iodide-chloride transporters.

At least in the case of the NTDT patient, the thyroid mutations also indicate a possible route of therapy via thyroid supplements for ameliorating any other issues of growth and development which in absence of this exome profiling may have remained attributed to the classical comorbidities of thalassaemia.

With reference to Table 6, where the homozygous mutations for the NTDT subject has been cited, no correlation could be drawn with the identified mutations and the current status of the patient.

The homozygous genetic mutation load was also higher and significant in the TDT individual than the NTDT subject.

This study identifies a need for a widespread genetic profiling of thalassaemia patients such that clinicians can better understand the genetic spectrum of common comorbidities of thalassaemia and be better equipped to identify and treat specific symptoms of thalassaemia with context of the whole.

Conflict of interest:

Authors declare there is no conflict of interest in the present study.

Acknowledgment:

The authors are grateful to Department of Biotechnology, Govt of India for funding this work [Sanc. No-BT/PR26461/MED/12/821/2018.Authors are

also express due acknowledgment to National Institute of Biomedical Genomics (NIBMG), Kalyani for providing necessary out sourced service for WES work]

References

1. Panja A, Chowdhury P, Chakraborty S, Ghosh TK, Basu A. Cross-Sectional Study for the Detection of Mutations in the Beta-Globin Gene Among Patients with Hemoglobinopathies in the Bengali Population. *Genet Test Mol Biomarkers*. 2017;21(1):39-45. doi:10.1089/gtmb.2016.0186
2. Setianingsih I, Williamson R, Daud D, Harahap A, Marzuki S, Forrest S. Phenotypic variability of Filipino beta(o)-thalassemia/HbE patients in Indonesia. *Am J Hematol*. 1999;62(1):7-12. doi:10.1002/(sici)1096-8652(199909)62:1<7::aid-ajh2>3.0.co;2-s
3. Sharma V, Kumar B, Kumar G, Saxena R. Alpha globin gene numbers: an important modifier of HbE/beta thalassemia. *Hematology*. 2009;14(5):297-300. doi:10.1179/102453309X446126
4. Charoenkwan P, Teerachaimahit P, Sanguansermsri T. The correlation of α-globin gene mutations and the XmnI polymorphism with clinical severity of Hb E/β-thalassemia. *Hemoglobin*. 2014;38(5):335-338. doi:10.3109/03630269.2014.952744
5. Bandyopadhyay S, Roychowdhury K, Chandra S, Das M, Dasgupta UB. Variable severity of beta-thalassemia patients of eastern India: effect of alpha-thalassemia and xmnI polymorphism. *Clin Exp Med*. 2001;1(3):155-159. doi:10.1007/s10238-001-8028-x
6. Phanrahan P, Yamsri S, Teawtrakul N, Fucharoen G, Sanchaisuriya K, Fucharoen S. Molecular Analysis of Non-Transfusion Dependent Thalassemia Associated with Hemoglobin E-β-Thalassemia Disease without α-Thalassemia. *Mediterr J Hematol Infect Dis*. 2019;11(1):e2019038. Published 2019 Jul 1. doi:10.4084/MJHID.2019.038
7. Ray R, Kalantri SA, Bhattacharjee S, et al. Association of alpha hemoglobin-stabilizing protein (AHSP) gene mutation and disease severity among HbE-beta thalassemia patients. *Ann Hematol*. 2019;98(8):1827-1834. doi:10.1007/s00277-019-03722-x
8. Oikonomidou, Paraskevi Rea, and Stefano Rivella. "What can we learn from ineffective erythropoiesis in thalassemia?." *Blood reviews* 32.2 (2018): 130-143.
9. Conran, Nicola. "Intravascular hemolysis: a disease mechanism not to be ignored." *Acta Haematol* 132.1 (2014): 97-99.
10. Gokhin, David S., and Velia M. Fowler. "Feisty filaments: actin dynamics in the red blood cell membrane skeleton." *Current opinion in hematology* 23.3 (2016): 206.
11. George, Alex, et al. "Altered phosphorylation of cytoskeleton proteins in sickle red blood cells: the role of protein kinase C, Rac GTPases, and reactive oxygen species." *Blood Cells, Molecules, and Diseases* 45.1 (2010): 41-45.
12. Konstantinidis, Diamantis G., et al. "Signaling and cytoskeletal requirements in erythroblast enucleation." *Blood, The Journal of the American Society of Hematology* 119.25 (2012): 6118-6127.
13. Lu, Yue, et al. "Toll-like receptors and inflammatory bowel disease." *Frontiers in immunology* 9 (2018): 72.
14. Monteiro, Juanita M., Catherine Harvey, and Giorgio Trinchieri. "Role of interleukin-12 in primary influenza virus infection." *Journal of virology* 72.6 (1998): 4825-4831.
15. Guo, Yifei, Wei Cao, and Ying Zhu. "Immunoregulatory Functions of the IL-12 Family of Cytokines in Antiviral Systems." *Viruses* 11.9 (2019): 772.
16. Cao, Anthony T., et al. "TLR4 regulates IFN-γ and IL-17 production

 doi:10.20944/preprints202103.0088.v1

by both thymic and induced Foxp3+ Tregs during intestinal inflammation." *Journal of leukocyte biology* 96.5 (2014): 895-905.

17. Xie, Peigen, et al. "Association of COL1A1 polymorphisms with osteoporosis: a meta-analysis of clinical studies." *International journal of clinical and experimental medicine* 8.9 (2015): 14764.
18. Jacobson, Eric M., Amanda Huber, and Yaron Tomer. "The HLA gene complex in thyroid autoimmunity: from epidemiology to etiology." *Journal of autoimmunity* 30.1-2 (2008): 58-62.
19. Shin, Dong-Hwan, et al. "HLA alleles, especially amino-acid signatures of HLA-DPB1, might contribute to the molecular pathogenesis of early-onset autoimmune thyroid disease." *PloS one* 14.5 (2019): e0216941.
20. Cangul, Hakan, et al. "Thyroid dyshormonogenesis is mainly caused by TPO mutations in consanguineous community." *Clinical endocrinology* 79.2 (2013): 275-281.
21. Peng and Meunier. Lewis System: Anti Lewis, Canadian Blood Services. Professional Education. Oct 2019

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2021 doi:10.20944/preprints202103.0088.v1

Table-1. Clinical findings of two subjects:

Clinical findings	Severe	Non severe
Present age	13 year	16 year
Sex	Female	Female
Age of onset	10 month	14 year
Baseline haemoglobin	6.1 gm/dl	7.8 gm/dl
Transfusion Interval	2 month	No transfusion yet
Spleen size	5 cm	9cm
MCV (fl)	61.1	64.4
MCH (pg)	18.1	19.3
RDW%	28.5	29.8
HPLC impression	HbE-β thalassemia	HbE-β thalassemia
HbA0 (%)	3.1	6.6
HbF (%)	50.3	31.5
HbA2+E (%)	42.6	57.4

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2021 doi:10.20944/preprints202103.0088.v1

Table 2: *Dene novo* variants of both TDT and NTDT subjects

Subject	Gene	db SNP ID	Chr	Ref	Alt	Func. refgene	Genotype
TDT	SPDYE1	rs573518672	chr7	AGG	-	splicing	het
	NBPF15	rs200795677	chr1	G	A	exonic	het
NTDT	BEX4	rs775129946	chrX	A	G	exonic	het
	SYNE2	rs546650178	chr14	C	T	exonic	het
	KCNH1	rs572710174	chr1	AAA	-	splicing	hom
	OR2A7	rs200085936	chr7	G	C	exonic	uncertain-het

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2021 doi:10.20944/preprints202103.0088.v1

Table 3: Involvement of the responsible alter Genes involved in different metabolic /or molecular pathway and their variants details for TDT as obtained through scheme 1 for minor allele hunting .

Pathway	Gene	db SNP ID	Chr	Ref	Alt	Func.refgene	Genotype
Adherens junction	FARP2	rs145392931	chr2	C	T	splicing	het
	LMO7	rs199685242	chr13	G	T	exonic	het
	ACTN3	rs200452235	chr11	G	A	exonic;splicing	het
	ACTN3	rs201487054	chr11	G	A	exonic	het
	FGFR1	rs542417198	chr8	A	G	splicing	het
	PTPRB	rs112571541	chr12	A	T	splicing	het
Protein digestion	COL1A1	rs66592376	chr17	G	T	splicing	het
	COL6A1	rs200261890	chr21	G	A	exonic	het
	COL6A3	rs146546544	chr2	G	A	exonic	het
	COL6A3	rs111228504	chr2	G	A	splicing	het
	COL21A1	rs528003403	chr6	A	G	splicing	het
IBD	RORA	rs200399670	chr15	A	T	exonic;splicing	het
	IL12RB1	rs369576406	chr19	C	T	exonic	het
	HLA-DOA	rs144931749	chr6	C	T	exonic	het
	TLR4	rs137853920	chr9	G	A	exonic	het
Amoebiasis	ACTN3	rs200452235	chr11	G	A	exonic;splicing	het
	ACTN3	rs201487054	chr11	G	A	exonic	het
	COL1A1	rs66592376	chr17	G	T	splicing	het
	C9	rs121909592	chr5	G	A	exonic	het
	MUC2	rs182669692	chr11	G	A	exonic	het
	TLR4	rs137853920	chr9	G	A	exonic	het
Glycolysis	ADH1C	rs6413444	chr4	G	A	exonic	het
	FBP1	rs200948424	chr9	C	G	splicing	het
	PCK2	rs201974284	chr14	G	A	exonic	het
	PGK2	rs140642109	chr6	T	A	exonic	het
PPAR signalling	CPT1A	rs374383052	chr11	C	T	exonic	het
	PLIN1	rs554749197	chr15	T	C	exonic	het
	PPARA	rs571930753	chr22	A	C	exonic	het
	PCK2	rs201974284	chr14	G	A	exonic	het

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2021 doi:10.20944/preprints202103.0088.v1

Table 4. Involvement of the responsible alter Genes involved in different metabolic /or molecular pathway and their variants details for TDT as obtained through scheme 2 for homozygous pathogenic loci hunting

Pathway	Gene	db SNP ID	Chr	Ref	Alt	Func. refgene	Genotype
Auto immune thyroid diseases	HLA-A	rs199474424	6	G	C	exonic	Homozygous
	HLA-A	rs2230991	6	G	A	exonic	Homozygous
	HLA-A	rs3173420	6	G	A	exonic	Homozygous
	HLA-B	rs1050723	chr6	G	A	exonic	hom
	HLA-B	rs1050570	chr6	T	C	exonic	hom
	HLA-B	rs1065386	chr6	G	C	exonic	hom
	HLA-B	rs1050529	chr6	C	T	exonic	hom
	HLA-C	rs41542414	chr6	A	T	exonic	hom
	HLA-DPB1	rs1042153	chr6	G	A	exonic	hom
	TPO	rs2175977	chr2	G	C	exonic	hom
Graft-versus-host disease/ Allograft rejection and Type I diabetes mellitus	HLA-A	rs199474424	chr6	G	C	exonic	hom
	HLA-A	rs2230991	chr6	G	A	exonic	hom
	HLA-A	rs3173420	chr6	G	A	exonic	hom
	HLA-B	rs1050723	chr6	G	A	exonic	hom
	HLA-B	rs1050570	chr6	T	C	exonic	hom
	HLA-B	rs1065386	chr6	G	C	exonic	hom
	HLA-B	rs1050529	chr6	C	T	exonic	hom
	HLA-C	rs41542414	chr6	A	T	exonic	hom
	HLA-DPB1	rs1042153	chr6	G	A	exonic	hom
Glycosphingolipid biosynthesis	FUT3	rs778986	chr19	A	G	exonic	hom
	FUT5	rs4807054	chr19	G	A	exonic	hom
	FUT9	rs3811069	chr6	A	G	exonic	hom
ECM-receptor interaction	COL6A6	rs9830253	chr3	G	A	exonic	hom
	ITGA11	rs4777035	chr15	G	A	exonic	hom
	LAMA5	rs944895	chr20	G	A	exonic;splicing	hom
	LAMA5	rs2427283	chr20	C	T	exonic	hom
	TNN	rs2072036	chr1	C	T	exonic	hom
	VTN	rs704	chr17	G	A	exonic	hom

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2021 doi:10.20944/preprints202103.0088.v1

Table 5: Involvement of the responsible alter Genes involved in different metabolic /or molecular pathway and their variants details for NTDT as obtained through scheme 1 for minor allele hunting .

KEGG pathway	**Gene**	**db SNP ID**	**Chr**	**Ref**	**Alt**	**Func.refgene**	**Gen**
Thyroid hormone synthesis	TG	rs35301433	chr8	A	G	exonic	het
	SLC26A4	rs375576481	chr7	A	T	exonic	het
	LRP2	rs41268685	chr2	C	T	exonic	het
	GPX6	rs562827573	chr6	C	T	exonic	het
Carbon metabolism	ACO1	rs554238047	chr9	G	T	exonic	het
	G6PD	rs5030868	chrX	G	A	exonic	het
	HK1	rs201626997	chr10	G	T	exonic;splicing	het
	OGDHL	rs146013158	chr10	C	A	splicing	het
	TKTL2	rs565930006	chr4	C	G	exonic	het

Table 6. Involvement of the responsible alter Genes involved in different metabolic /or molecular pathway and their variants details for NTDT as obtained through scheme 2 for homozygous pathogenic loci hunting

Pathway	**Gene**	**db SNP ID**	**Chr**	**Ref**	**Alt**	**Func.refgene**	**Genotype**
Protein digestion & absorption	CELA3B	rs7528405	chr1	C	T	exonic;splicing	hom
	COL6A6	rs9830253	chr3	G	A	exonic	hom
	COL6A6	rs61629992	chr3	C	T	exonic	hom
	COL11A1	rs3753841	chr1	G	A	exonic	hom
	COL12A1	rs970547	chr6	C	T	exonic;splicing	hom
	COL14A1	rs4870723	chr8	A	C	exonic	hom
	COL24A1	rs11161732	chr1	G	A	exonic	hom

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2021 doi:10.20944/preprints202103.0088.v1

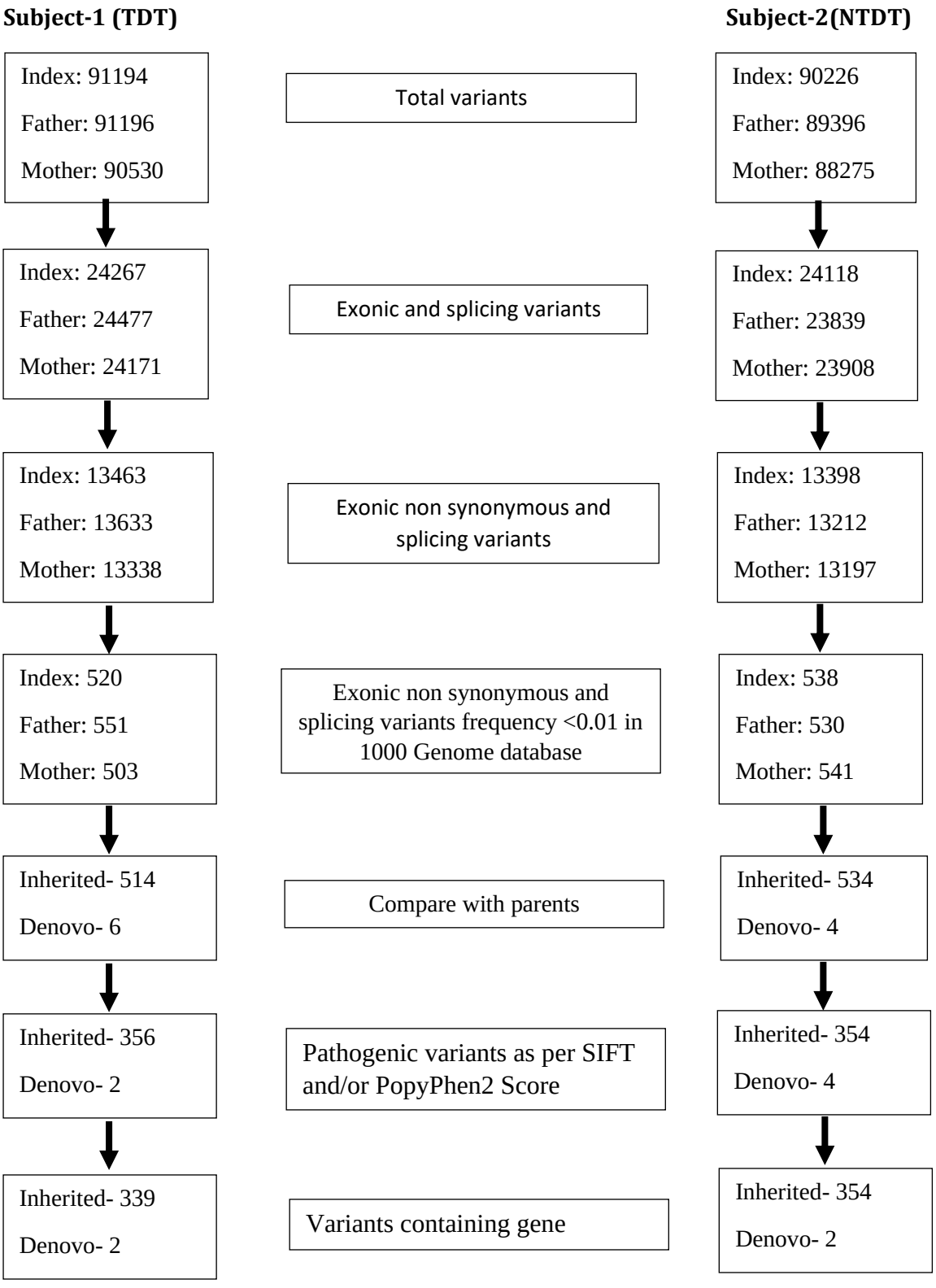

Fig 1: Flowchart of the Scheme 1-for finding of the different inherited variants with minor allele frequency

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2021 doi:10.20944/preprints202103.0088.v1

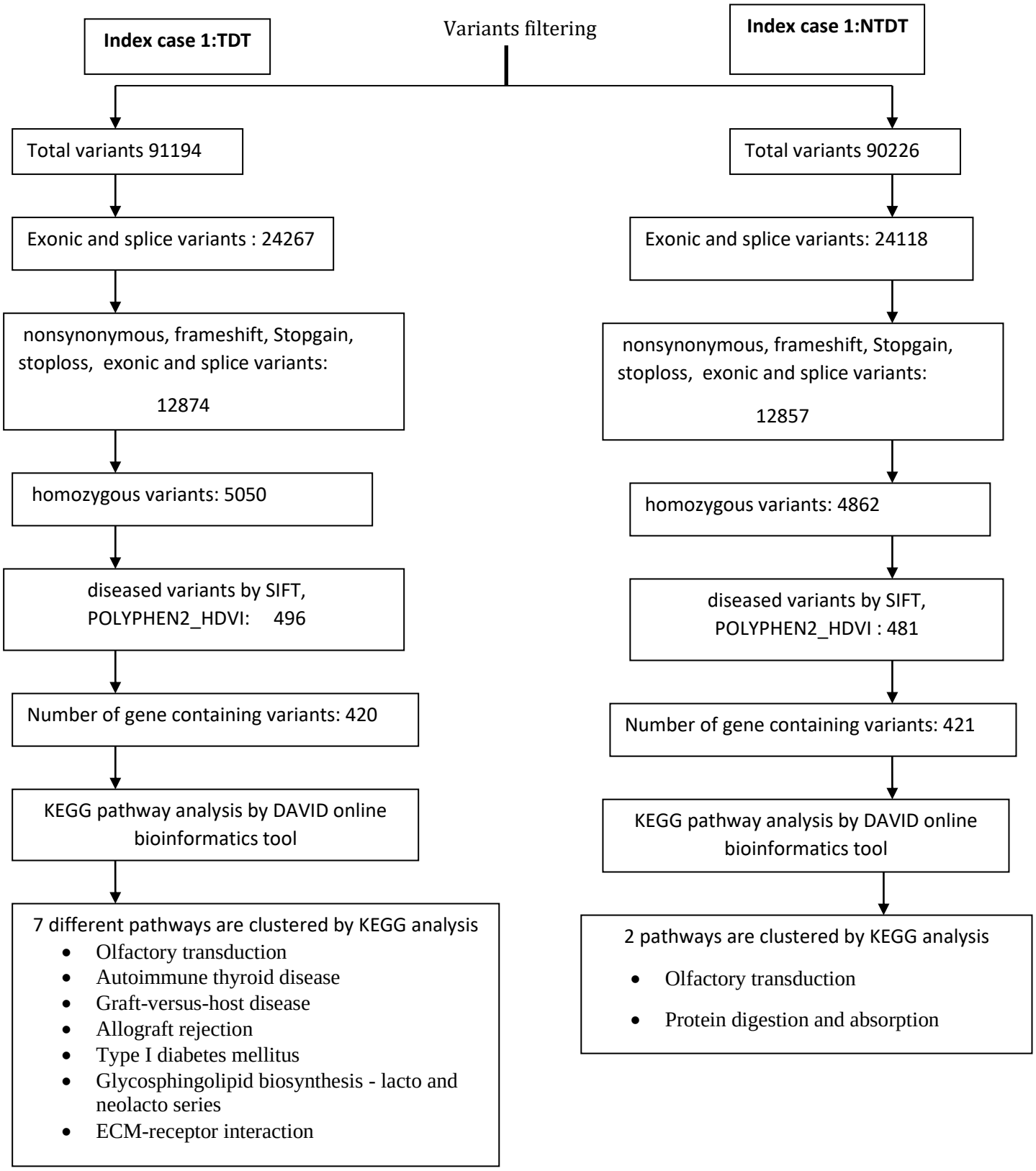

Fig 2: Flowchart of the Scheme 2-for finding of the different inherited homozygous pathogenic variants and respective gene functional clustering using KEGG analysis

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2021 doi:10.20944/preprints202103.0088.v1

Supplementary Information

Supplementary Table 1: Functional annotation of the gene out of de novo variants of both the TDT and NTDT index subjects.

Subject	Gene	Gene Function
TDT	SPDYE1	Interacting selectively and non-covalently with a protein kinase, any enzyme that catalyzes the transfer of a phosphate group, usually from ATP, to a protein substrate. Diseases associated with SPDYE1 include Williams-Beuren Syndrome
	NBPF15	This gene is a member of the neuroblastoma breakpoint family (NBPF), Members of this gene family are characterized by tandemly repeated copies of DUF1220 protein domains. Diseases associated with NBPF15 include Neuroblastoma and Hydrolethalus Syndrome
NTDT	BEX4	May play a role in microtubule deacetylation by negatively regulating the SIRT2 deacetylase activity toward alpha-tubulin and thereby participate in the control of cell cycle progression and genomic stability. The proteins encoded by some of the other members of this family act as transcription elongation factors which allow RNA polymerase II to escape pausing during elongation.
	SYNE2	Multi-isomeric modular protein which forms a linking network between organelles and the actin cytoskeleton to maintain the subcellular spatial organization. As a component of the LINC (LInker of Nucleoskeleton and Cytoskeleton) complex involved in the connection between the nuclear lamina and the cytoskeleton. play an important role in the transmission of mechanical forces across the nuclear envelope and in nuclear movement and positioning.
	KCNH1	Voltage-gated potassium (Kv) channels represent the most complex class of voltage-gated ion channels from both functional and structural standpoints. Their diverse functions include regulating neurotransmitter release, heart rate, insulin secretion, neuronal excitability, epithelial electrolyte transport, smooth muscle contraction, and cell volume
	OR2A7	OR2A7 (Olfactory Receptor Family 2 Subfamily A Member 7) is a Protein Coding gene. Among its related pathways are Signaling by GPCR and Olfactory transduction.