OPEN ACCESS Freely available online

Journal of Proteomics & Bioinformatics

Research Article

The Unmasking of 'Mitochondrial Adam' and Structural Variants Larger than Point Mutations as Stronger Candidates for Traits, Disease Phenotype and Sex Determination

Abhishek Narain Singh*

Schiller International University, Heidelberg Campus, Zollhofgarten 1, 69115 Heidelberg, Germany

ABSTRACT

Background: Structural Variations, SVs, in a genome can be linked to a disease or characteristic phenotype. The variations come in many types and it is a challenge, not only determining the variations accurately, but also conducting the downstream statistical and analytical procedure.

Method: Structural variations, SVs, with size 1 base-pair to 1000s of base-pairs with their precise breakpoints and single-nucleotide polymorphisms, SNPs, were determined for members of a family. The genome was assembled using optimal metrics of ABySS and SOAP de novo assembly tools using paired-end DNA sequence.

Results: An interesting discovery was the mitochondrial DNA could have paternal leakage of inheritance or that the mutations could be high from maternal inheritance. It is also discovered that the mitochondrial DNA is less prone to SVs re-arrangements than SNPs, which propose better standards for determining ancestry and divergence between races and species over a long-time frame. Sex determination of an individual is found to be strongly confirmed using calls of nucleotide bases of SVs to the Y chromosome, more strongly determined than SNPs. We note that in general there is a larger variance (and thus the standard deviation) in the sum of SVs nucleotide compared to sum of SNPs of an individual when compared to reference sequence, and thus SVs serve as a stronger means to characterize an individual for a given trait or phenotype or to determine sex. The SVs and SNPs in HLA loci would also serve as a medical transformational method for determining the success of an organ transplant for a patient, and predisposition to diseases apriori. The sample anonymous dataset shows how the *de-novo* mutation can lead to non-inherited disease risk apart from those which are known to have a disease to mutation association. It is also observed that mtDNA is highly subjected to mutation and thus the factor for a lot of associated maternally inherited diseases.

Conclusion: 'Mitochondrial Adam' can be a fair reality as certainly the biparental mode of mtDNA puts in question the theory of 'mitochondrial Eve'. SVs would serve as a stronger fingerprint of an individual contributing to his traits, sex determination, and drug responses than SNPs.

Keywords: Bioinformatics; High-performance computing; Medical informatics; Genetics; Genomics; NGS

Abbreviations: SVs: Structural Variations; InDels: Insertions and Deletions; DIPs: Deletion and Insertion Polymorphism; SNPs: Single Nucleotide Polymorphisms; mtDNA: mitochondrial Deoxyribonucleic Acid; CNV: Copy Number Variations; SNVs: Single Nucleotide Variations

BACKGROUND

Customization of the genome and biological material analysis for a tailor-made solution individual specific is the next bio and medical selling proposition. The high false discovery rate of structural variation algorithms, even in deeply sequenced individual genomes of the order of 30x average coverage [1,2] suggests that for lower coverage the larger problem will to get rid of false positives. Nevertheless, the results with coverage as low

Correspondence to: Abhishek Narain Singh, Schiller International University, Heidelberg Campus, Zollhofgarten 1, 69115 Heidelberg, Germany, E-mail: abhishek.narain@iitdalumni.com

Received: September 22, 2021; **Accepted:** October 06, 2021; **Published:** October 13, 2021

Citation: Singh AN (2021) The Unmasking of 'Mitochondrial Adam' and Structural Variants Larger than Point Mutations as Stronger Candidates for Traits, Disease Phenotype and Sex Determination. J Proteomics Bioinform.14:557.

as 3-5x also could have meaningful findings and be deployed for several genomes analysis, which would make sense on a population-wide scale at a relatively lower cost, such as in the 1000 genomes project, phase I [3]. An arXiv preprint of this paper was published in early 2021 [4] and the abstract was published in the book of abstract of BIOCOMP 2020 [5].

In the current article, we discuss and extend the findings of mtDNA and Y-chromosome characteristics in terms of SNPs and SVs which was first presented in the year 2012 at a personalized genomics conference [6], followed by an ex-tended article in 2013 at the international BIOCOMP conference [7], and it would be highly unlikely that many such similar experiments would not converge to some discoveries as it did happen in later years [8-10]. This paper serves as an extended paper of [11,12], particularly adding to them the clinical procedure following the variants extraction, and emphasis and clarity on mitochondrial inheritance possibility through paternal lines and thus the associated diseases. Most recently Luo et al. [13] in 2018, also came up with a large cohort that showed it is possible to have mtDNA being biparentally inherited, a discovery which was already made 6 years ago by Singh AN [6] in the year 2012, although a rare case of paternal mtDNA inheritance was also reported earlier [14]. The article thus has now been appropriately titled as mitochondrial Adam, to give due weightage of possibility of paternal or bi-paternal inheritace of mitochondrial DNA as previously mitochondrial Eve [15] paper discussed the possibility of maternal inheritance of mitochondria as the only possibility. In addition, the paper also highlights the rising importance of mutations larger than just the point mutations, as a significant contributing factor to determine end phenotype and trait.

With regards to discovery of SVs association to traits and phenotype of individual, here, more precisely we examine the InDels in SVs which are also known as DIPs. The analysis is then done for the whole family for matches of the SNPs variations to known pathogenic SNPs database, ClinVar [16,17] for the sake of completion and way forward for medical diagnostics and was also presented at the international BIOCOMP conference [18] in July 2019.

Variations at specific loci in the genome have been associated with recurrent genomic rearrangements as well as with a variety of diseases, including color blindness, psoriasis, HIV susceptibility, Crohn's disease and lupus glomerulonephritis [19-24]. Figure 1 summarizes in a broad sense the various variations that can occur in a genome in comparison to a reference genome, as was initially published [25]. This only enhances the importance of a comprehensive catalog for genotype and phenotype correlation studies when the rare or multiple variations in gene underlie characteristic or disease susceptibility [26,27]. Microarrays [28-30] and sequencing [31-34] reveal that the Structural Variants (SVs) contribution is significant in characterizing populations [35] and disease [36] characteristics. Interestingly the HLA domain in chromosome VI of an individual, which is the MHC region in humans, would be interesting in being decoded for the variations, as a lesser difference between two individuals could imply a greater likely of success in an organ transplant. With time the sequencing of human genomes now be-come routine, the spectrum of structural variants and Copy Number Variants (CNVs) has widened to include much smaller events. The important aspect now is to know how genomes vary at large as well as fine scales and by what magnitude does it impact a population in general and an individual. There are databases such as OMIM [37], ClinVar, dbSNP [38], PharmGKB [39], HGVbaseG2P [40] now named GWAS Central, UniProt [41], etc., where genotype to phenotype associations are maintained and regularly updated. There have been several new tools made available which can detect variations without the need for assembling the genome for the individual, such as those used in the 1000 Genome Project consortium which finds great applicability in case the coverage of sequences is low [1] and has, so to speak, yet have a profound impact at a population level. In this article, we share the results of the variations detected in a family of four individuals' viz., father, mother, and two daughters.

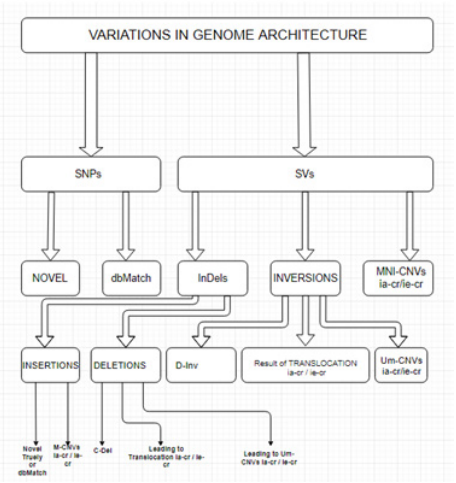


Figure 1: Variations in genome architecture. Um: Un matching; M: Matching; MNI: Matching Non Insertion; ia-cr: intra chromosomal; ie-cr: inter chromosomal; SNPs: Single Nucleotide Polymorphism; SNV: Single Nucleotide Variations; InDels: Insertions and Deletions; CNVs: Copy Number Variations; Translocation: Single copy match elsewhere in genome. Tandem Duplication and Multiplication is a subset of various types of CNVs. Mobile element insertion is a subset of M-CNVs, C-Del: Complete Deletion; D-Inv: Direct Inversion. The classification is done based on the characteristic of variation and not on basis of sizes of variations. Figure plotted using www.draw.io.

METHOD

The blood samples of a family were collected in Amsterdam, although they might not be individuals who are of direct Dutch descent as Amsterdam is a cosmopolitan city. To keep their identities anonymous, we will refer to them as A105A, A105B, A105C, and A105D, respectively. The DNA was extracted and sequenced on Illumina HiSeq sequencer with an average coverage of more than 12x across the genome and with the raw read length of 90 bases at either end of the paired end reads with an average insert size of about 470 bases. As there are many copies of mitochondrial DNA in a cell, the sequencing coverage of mitochondrial DNA would be several folds higher than 12x. The reads were then assembled into respective contigs using parallel assembler ABySS [42] version 1.3.1 with optimal parameters of kmer size (k) of 49 and minimum reads to make a consensus contigs (n) of 3 to yield highest possible N50 value

for the contigs ~3000. Default values of SOAP *de-novo* [43] were also used for assembling the genomes. SSPACE [44] was also used for assembly. On average it required about 140 GB of RAM in a shared environment and 49 computing wall-clock hours on a symmetric multiprocessor cluster with 6 computing cores each of capacity 2.6 GHz. The assemblies of the four individuals were then aligned globally to the NCBI human reference genome, Build 37, followed by extraction of SVs information of category insertions and deletions only (InDels), and single nucleotide polymorphisms (SNPs) on regions of misalignment [45,46]. Genome plot for the A105 A, B, C and D is shown in Figures 2a and 2b where one can graphically obtain estimates of regions of alignments and misalignments with the reference genome NCBI HuRef build 37.

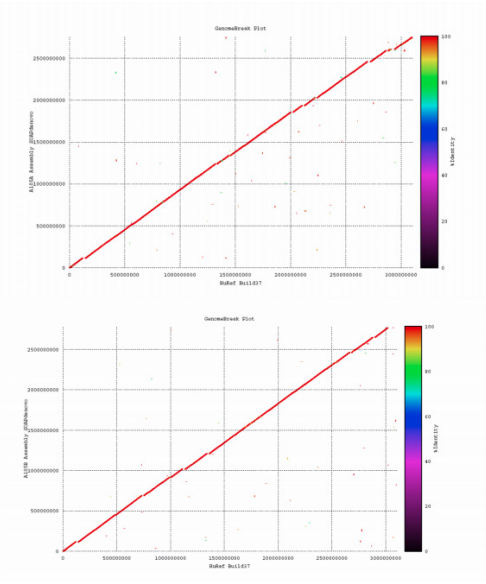


Figure 2a: A105A, A105B. The large breaks and differences between the graphs are visible to naked eye, though much of the variations can be found when zoomed in.

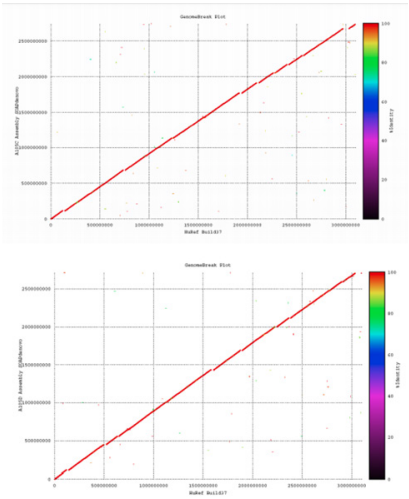


Figure 2b: A105C, A105D. The large breaks and differences between the graphs are visible to naked eye, though much of the variations can be found when zoomed in.

The total time for the alignment and extraction of information on a single computing core of 2.6 GHz capacity came out to about 85 wall-clock hours, for each assembly.

RESULTS

Scientific outcomes

The SNPs of the whole genome can also be used for generic ancestry and divergence determination. For this reason, we also conducted a SNP genome-wide inheritance analysis and found that child A105C and child A105D had about 20.82% and 21.02%, respectively, of its SNPs that were novel compared to those present in the parents i.e. A105A and A105B (R code for analysis provided in supplementary). The consistency in both the children of about 21% SNPs being different from those found in either of the parents gives confidence in the methodology deployed. The difference could also be attributed to the aspect of only 14x genome average coverage and thus less robustness in detecting sequencing errors. Below in Figure 3 is the plot of the sum of the bases of InDels and SNPs for chromosome 20 of the A105 family. The sum of the bases of InDels have in-creased in the children when compared to their parents while the levels of SNPs remain more-or-less the same, clearly speaking of higher variance and thus higher standard deviation from mean in the InDels compared to the SNPs. This interesting observation of higher standard deviation in the InDels compared to SNPs, even in just one generation, clearly points out that InDels would be a stronger candidate for attributing a genetic signature of an individual and can be thus linked more confidently with determining the sex of an individual and associating with traits and disease phenotypes.

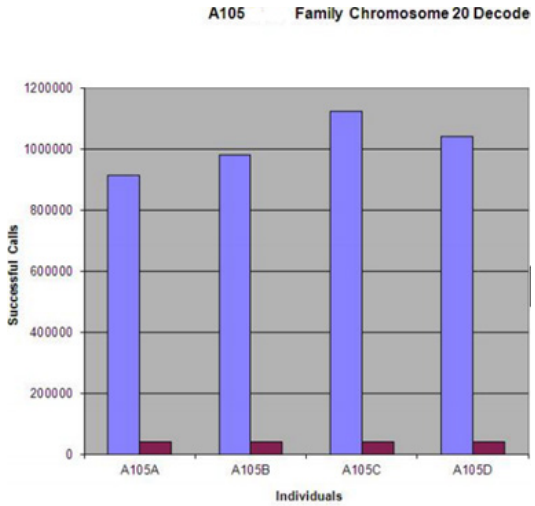


Figure 3: SVs and SNPs from only Chr 20.

Note: ( ) Assembly InDels+SNPs, ( ) Assembly SNPs

Figure 4 is a plot of the sum of the bases of SVs and SNPs respectively for the whole genome determined for the A105 family. The corresponding values for the plot along with mean and standard deviation values for SVs and SNPs for the family is shown in Table 1, here again, we notice that the children

have a relatively higher number of bases for SVs than their parents, though the levels of SNPs remain more-or-less the same. This finding thus proposes that even in one generation of the offspring, there can be a significant rearrangement in the genetic background to produce greater variance (and thus standard deviation) in the variation of genotype and thereby having an effect on phenotypes, and that the children are not an exact clone of the set of chromosomes they inherit from either parent as there will be significant variation, even when simply compared to the chromosomes of the parents that they inherited. The changes in SNPs are more restricted than insertions or deletions, and thus SVs serve as a stronger mean as a fingerprint and characteristics of an individual when analysed genome wide.

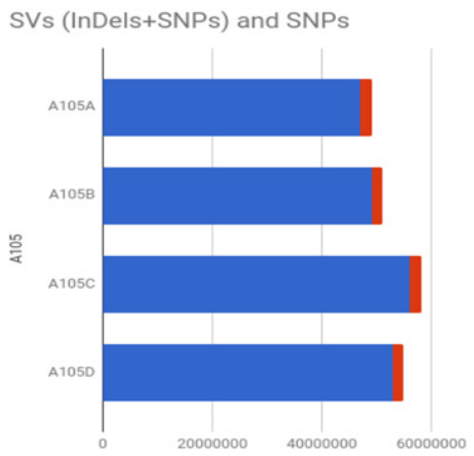


Figure 4: Genome Wide SVs and SNPs: SNPs are more conserved genome wide. SVs: Structural Variations; SNP's: Single Nucleotide Polymorphism.

Note: ( ) SVs(InDels+SNPs), ( ) SNPs

Table 1: The standard deviation of SNPs (Single Nucleotide Polymorphism) being far smaller than that of SVs (Structural variations) for whole genome.

A105	SVs (In Dels+SNPs)	SNPs	Statistics	Value
A105A	47083259	2001074	SD SVs	4018934.7
A105B	49139752	2022063	SD SNPs	24543.9
A105C	56138900	2021242	Mean SVs	51318148.3
A105D	52910682	2059879	Mean SNPs	2026064.5

In Figure 5 and Figure 6 you will see that the calls of bases on the Y chromo-some of InDels and SNPs, respectively, is far higher for the father than the mother or the two daughters, thereby clearly being able to differentiate male from female. The pseudoautosomal regions, PAR1, PAR2, are homologous sequences of nucleotides on the X and Y chromosomes [47] where Genetic recombination (occurring during sexual reproduction) is known to be limited only to the pseudoautosomal regions (PAR1 and PAR2) of the X and Y chromosomes. Thus, small matches will be expected to be found in the female candidates when making calls for SVs and SNPs having the reference Y-chromosome from NCBI Build 37 as the counter alignment pair, as observed in our data and plotted in Figures 5 and 6. It is also observed that the difference in the calls of sums of bases for InDels is far higher than the calls for the sum of the bases for SNPs, thereby proposing that the former is a stronger means to determine the sex of an individual than the latter. This also proposes that contrary to what is observed genome-wide, the SVs have higher selection pressure than the SNPs in the Y-chromosome. This strong selection pressure of Y-chromosome associated SVs compared to SNPs in woman, as also shown in Table 2 where the SVs comprise only about 0.44% of total SVs in male while the SNPs comprise about 23% of that number of male, which would usually be those of PAR1 and PAR2 region, proposes a stronger method by SVs to determine paternal inheritance in a woman, by the data illustrated here.

Even though this work involved had about 14x on average coverage for the genome, since the mitochondrial DNA comprises anywhere from 2 to 10 copies of the DNA for each mitochondrion [48], and since each cell in itself comprises of 1000s of mitochondria, the effective coverage for a mitochondrial genome would be in the order of 14x × Mean (2,10) × 1000, which is Order(14x × 6 × 1000), which is about Order(84,000x). With such a heavy coverage of the mtDNA, it is highly unlikely that the assembly process of the mtDNA genome would be faulty given that most assembly work typically ranges for genome coverage in the median value of 49x with a population size of 27 as per Ekblom and Wolf [49]. From the already existing knowledge of inheritance of mitochondrial DNA, one would expect all the SNPs and SVs successful calls in mother to be found in all the children as well, as mitochondrial DNA is known to be maternally inherited. This is because mitochondrial DNA material is present in the cytosol of a cell and not in the nucleus, and there is a lesser possibility for the cytosol of the sperm cells to integrate with the cytosol of the mother ova and is known to be destroyed at fertilization. So, for determining maternal inheritance, mtDNA is the same as his mother's mtDNA, which is the same as her mother's mtDNA and so on.

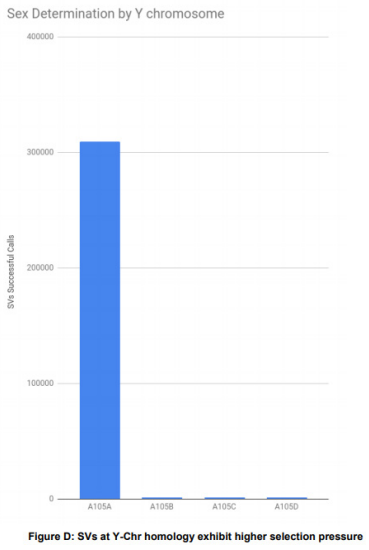


Figure 5: SVs at Y-Chr homology exhibit higher selection pressure. Svs: Structual Variations. Y-Chr: Y Chromosome.

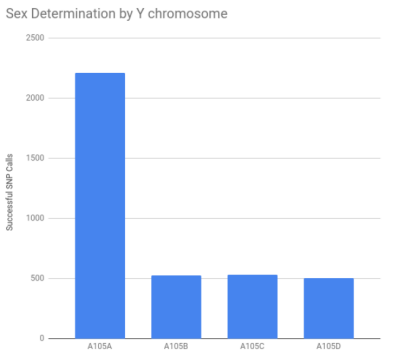


Figure 6: SNPs at Y-Chr homology exhibit lower selection pressure, yet good enough to distinguish gender. SNPs: Single Nucleotide Polymorphism

Table 2: SVs (Structural Variations) highly conserved in terms of number for Y-chromosome with mean 0.45% than SNPs 23.57%. Percentage is calculated keeping the number for A105A (father) as the reference.

A105 Chr Y	SVs (InDels+SNPs)	SNPs (Single Nucleotide Polymorphism)	%age representation SVs	%age representation SNPs (Single Nucleotide Polymorphism)
A105A	309158	2208	100	100
A105B	1376	525	0.445079862	23.77717391
A105C	1600	534	0.517534723	24.18478261
A105D	1195	502	0.386533747	22.73550725
		Average	0.449716111	23.56582126

Our findings for A105 family analysis revealed contradicting results. Not all SNPs and SVs present in the mother were found to be present in the children. There were cases found where a SNP was found to be present in the father and a child but not in mother. Table 3 shows the list of SNPs and Table 4 shows the list of SVs in A105 family. This proposes a discovery that mitochondrial DNA can have paternal sources of inheritance as well, though they can also be a result of *de-novo* genetic changes rather than inheritance. Further, comparing Table 3 and Table 4, it is discovered by the observation that mitochondrial DNA is less prone to SVs than SNPs, and that can be possibly attributed to the fact that mitochondrial DNA is not exposed to the phenomenon of crossing-over of genetic material as is the case with chromosomes. This proposes a discovery that mitochondrial DNA can have paternal or mixed paternal-maternal sources of inheritance as well, though most likely they can be a result of *de-novo* genetic changes rather than inheritance. If *de-novo* mutations happen to be the case, it would again oppose the 'mtDNA bottleneck' theory which states that mtDNA is subjected to high bottleneck. The results by Table 3 and Table 4 would then propose a 'mtDNA bottleneck leakage' theory, where possibly one of the factors that were proposed in ("Two ways to make") [50] has caused a bottleneck for mutation that must have gone uncontrolled. The heteroplasmy nature of mtDNA cannot be ruled out either, but the fact that maternal mtDNA SNPs are different than each of the offsprings raises several questions. Further, the ratio of SVs bases calls to the size of genome is significantly less for mitochondrial DNA (of the order of 2.35×10^{-4}) than for the whole genome (of the order of 2.5×10^{-3}), thereby providing further evidence that structural variations in mitochondrial DNA has higher selection pressure than the rest of the genome and is thus a more rare event in the mitochondria relative to the rest of the genome. This ratio remains comparable to the rest of the genome when considered for SNPs (of the order of 5.3×10^{-4} for mitochondria and the order of 8.0×10^{-4} for whole-genome).

Mitochondrial DNA and Y-chromosome DNA have been widely used to determine maternal and paternal ancestry respectively, such as in recent findings for Native Americans and Indigenous Altaians [51]. Based on the discoveries above, it can thus be safely concluded that if we continue with ancestry determination by mitochondrial DNA, then SVs would serve as better means to determine ancestry for a longer period than SNPs, as they are relatively more rare events. At the same time, the SNPs of the mitochondria would serve as a better candidate for the characteristic signature of the individual and can be used to determine ancestry and divergence for a relatively shorter period. Having said that, it would still be pro-posed that given that there is the possibility of mitochondrial DNA to be inherited by the father as well; maternal ancestry determination by mtDNA should be re-phrased as simply ancestry determination by mtDNA. This will also mean that all the analysis, which different scientists across the globe have been conducting so far assuming mtDNA to be maternally inherited, will need a complete change in the understanding and knowledge generated. As it is confirmed that the Y-chromosome is completely paternally inherited, ancestry determination by 'Y line tests' as Y-chromosomes are confirmed to be inherited from the father remain a good methodology. Further, as observed and stated above, since SVs have higher selection pressure than SNPs for the Y-chromosome, the SVs will serve as a better means for paternal ancestry determination for a relatively longer time-span and the SNPs would serve as a better candidate to determine paternal ancestry and divergence in a relatively shorter time-span.

Table 3: Genome breaks findings: SNP of mother mtDNA not inherited by a female offspring, but found similar to that of father mtDNA. 331 is the position and A stands for male member. 16519 is the position and T stands for Thymine. 16527 is the position and C stands for Cytosine.

Genome break				Plot data			
Stringent parameters				Lenient parameters			
A105A	A105B	A105C	A105D	A105A	A105B	A105C	A105D
331 A	331 A	131 T	339 A	331 A	331 A	131 T	339 A
493 A	1476 G	750 A	6474 A	493 A	15380 A	750 A	6474 A
16496 G	1518 C	4769 A	6497 T	1476 G	15408 A	4769 A	6497 T

16519 T	15380 A	16519 T	15476 C	16519 T	16220 A	16519 T	15476 C
16527 C	15408 A			16496 G	16249 T		
	16220 A			1518 C	16437 T		
	16249 T			16527 C	16469 T		
	16437 T						
	16469 T						

Table 4: Genome break findings-a SV of mother MTDNA (Mitochondrial DNA) not present in a female offspring. A105 is the family and A stands for the male member. 3107 is the position of the nucleotide and N stands for the nucleotide, 314 is the position and C stands for Cytosine, 523 is the position and A stands for Adenosine.

Genome break				Plot data			
Stringent parameters				Lenient parameters			
A105A	A105B	A105C	A105D	A105A	A105B	A105C	A105D
3107 N	3107 N	3107 N	3107 N	3107 N	3107 N	3107 N	3107 N
314 C	314 C	Missing	314 C	314 C	314 C	Missing	314 C
522 C		4824 N		522 C			
523 A				523 A			

Clinical applications

If the immunologic responses after the grafting of an organ from a donor to the receptor are known before conducting the transplant, we can be more predictive of the chances of success of the transplantation. The immunologic responses are dictated by the MHC region of the genome, which in humans corresponds to the HLA domain in chromosome VI.

DNA editing technologies such CRISPR/Cas9 [52], in the hope that one day the method would be even more precise, combined with this kind of study, as mentioned in this paper for detecting any unwanted structural variations, can be a great promise for the future. For this purpose, we first matched the SNPs from the A105 family members to the known OMIM database and stored the result in Microsoft Excel sheet as shown in Figure 7. From this table, filtering out those entries that have clinical association is straightforward, as shown in Figure 8. Databases such as dbSNP version 130 was also used for the very purpose and to ClinVar database version 20140211. Figures 9-12 show the results of ClinVar analysis for A105 A, B, C, and D having 26, 25, 31 and 24 entries respectively. From this data, we could see that apart from SNPs of clinical significance that was inherited from the parents, both children A105C and A105D also had 7 and 5 novel clinically relevant SNPs due to *de-novo* mutation (calculation in R script in supplementary). Figures 13 and 14 show the novel SNPs.

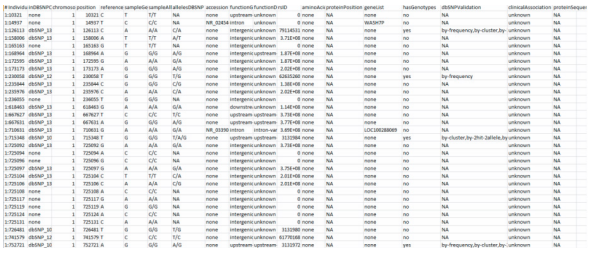


Figure 7: Snippet of SNP OMIM database match. SNP: Single Nucleotide Polymorphism; OMIM: Online Mendelian Inheritance in Man.

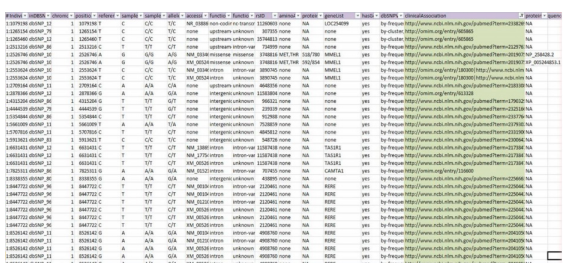
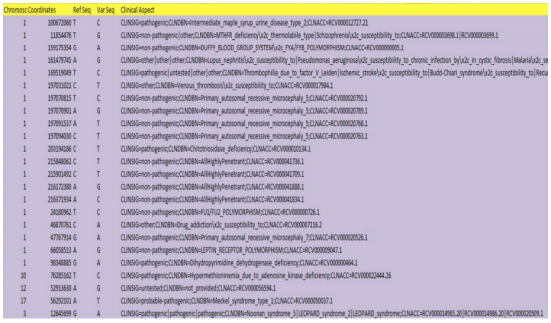


Figure 8: Snippet of Clinically relevant SNPs (Single Nucleotide Polymorphism) using OMIM (Online Mendelian Inheritance in Man).

Chromosome	Co-ordinates	Ref Bas	Variant	Clinical Aspect
1	100672060	T	C	CLINSIG=pathogenic;CLNDBN=Intermediate_maple_syrup_urine_disease_type_2C;CLNACC=RCV000012727.21
1	114377568	A	G	CLINSIG=other|other|other|other;CLNDBN=Diabetes_mellitus\x2c_insulin-dependent\x2c_susceptibility_to|Rheumatoid_arthritis|
1	169519049	T	C	CLINSIG=pathogenic|untested|other|other;CLNDBN=Thrombophilia_due_to_factor_V_Leiden|Ischemic_stroke\x2c_susceptibility_to|Bud
1	193218843	C	G	CLINSIG=non-pathogenic;CLNDBN=Parathyroid_carcinoma;CLNACC=RCV000020880.2
1	197031021	C	T	CLINSIG=other;CLNDBN=Venous_thrombosis\x2c_susceptibility_to;CLNACC=RCV000017984.1
1	197070815	T	C	CLINSIG=non-pathogenic;CLNDBN=Primary_autosomal_recessive_microcephaly_5;CLNACC=RCV000020792.1
1	197070901	A	G	CLINSIG=non-pathogenic;CLNDBN=Primary_autosomal_recessive_microcephaly_5;CLNACC=RCV000020789.1
1	197091537	A	T	CLINSIG=non-pathogenic;CLNDBN=Primary_autosomal_recessive_microcephaly_5;CLNACC=RCV000020768.1
1	197094030	C	T	CLINSIG=non-pathogenic;CLNDBN=Primary_autosomal_recessive_microcephaly_5;CLNACC=RCV000020763.1
1	203194186	C	T	CLINSIG=pathogenic;CLNDBN=Chitotriosidase_deficiency;CLNACC=RCV000010134.1
1	216172380	A	G	CLINSIG=non-pathogenic;CLNDBN=AllHighlyPenetrant;CLNACC=RCV000041888.1
1	216173934	A	C	CLINSIG=non-pathogenic;CLNDBN=AllHighlyPenetrant;CLNACC=RCV000041834.1
1	216595306	C	T	CLINSIG=non-pathogenic;CLNDBN=AllHighlyPenetrant;CLNACC=RCV000041833.1
1	223285200	G	A	CLINSIG=other|pathogenic|pathogenic;CLNDBN=Legionellosis|Systemic_lupus_erythematosus\x2c_resistance_to\x2c_1|Melioidosis\x2c_
1	46870761	C	A	CLINSIG=other;CLNDBN=Drug_addiction\x2c_susceptibility_to;CLNACC=RCV000007116.2
1	47767914	G	A	CLINSIG=non-pathogenic;CLNDBN=Primary_autosomal_recessive_microcephaly_7;CLNACC=RCV000020526.1
1	55518316	C	T	CLINSIG=non-pathogenic;CLNDBN=Familial_hypercholesterolemia;CLNACC=RCV000030351.1
1	63872032	T	C	CLINSIG=pathogenic|non-pathogenic;CLNDBN=Congenital_disorder_of_glycosylation_type_1C|AllHighlyPenetrant;CLNACC=RCV000023375
1	66058513	A	G	CLINSIG=non-pathogenic;CLNDBN=LEPTIN_RECEPTOR_POLYMORPHISM;CLNACC=RCV000009047.1
1	66075952	G	C	CLINSIG=non-pathogenic;CLNDBN=LEPTIN_RECEPTOR_POLYMORPHISM;CLNACC=RCV000009050.1
1	98348885	G	A	CLINSIG=pathogenic;CLNDBN=Dihydropyrimidine_dehydrogenase_deficiency;CLNACC=RCV000000464.1
19	57328825	C	T	CLINSIG=untested;CLNDBN=Malignant_melanoma;CLNACC=RCV000083653.2
3	12645699	G	A	CLINSIG=pathogenic|pathogenic|pathogenic;CLNDBN=Noonan_syndrome_5|LEOPARD_syndrome_2|LEOPARD_syndrome;CLNACC=RCV000
4	2833309	T	A	CLINSIG=pathogenic;CLNDBN=Fibrous_dysplasia_of_jaw,Fibrous_dysplasia_of_jaw,Fibrous_dysplasia_of_jaw;CLNACC=RCV000007985.1,RCV
5	112130448	G	A	CLINSIG=other;CLNDBN=Familial_colorectal_cancer;CLNACC=RCV000074005.1
9	71836364	G	T	CLINSIG=unknown;CLNDBN=AllHighlyPenetrant;CLNACC=RCV000037077.1

Figure 9: ClinVar matches A105A.

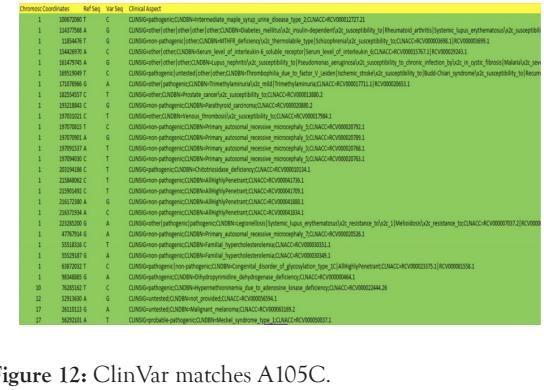


Chromosome	Coordinates	Ref Seq	Var Seq	Clinical Aspect
1	100672060	T	C	CLINSIG=pathogenic;CLNDBN=Intermediate_maple_syrup_urine_disease_type_2C;CLNACC=RCV000012727.21
1	114377568	A	G	CLINSIG=non-pathogenic|other|other|other;CLNDBN=Diabetes_mellitus\x2c_insulin-dependent\x2c_susceptibility_to|Rheumatoid_arthritis|Systemic_lupus_erythematosus\x2c_susceptibility_to;CLNACC=RCV000003698.1|RCV000003699.1
1	11854476	T	G	CLINSIG=non-pathogenic|other|other|other;CLNDBN=MTHFR_deficiency\x2c_thermolabile_type|Schizophrenia\x2c_susceptibility_to;CLNACC=RCV000003698.1|RCV000003699.1
1	155175314	G	A	CLINSIG=non-pathogenic;CLNDBN=DUFFY_BLOOD_GROUP_SYSTEM\x2c_FY4/FY8_POLYMORPHISM;CLNACC=RCV000000005.1
1	161479745	A	G	CLINSIG=other|other|other;CLNDBN=Lupus_nephritis\x2c_susceptibility_to|Pseudomonas_aeruginosa\x2c_susceptibility_to_chronic_infection_by\x2c_in_cystic_fibrosis|Malaria\x2c_se
1	169519049	T	C	CLINSIG=pathogenic|untested|other|other;CLNDBN=Thrombophilia_due_to_factor_V_Leiden|Ischemic_stroke\x2c_susceptibility_to|Budd-Chiari_syndrome\x2c_susceptibility_to|Recu
1	170331021	C	T	CLINSIG=other;CLNDBN=Venous_thrombosis\x2c_susceptibility_to;CLNACC=RCV000017984.1
1	197070815	T	C	CLINSIG=non-pathogenic;CLNDBN=Primary_autosomal_recessive_microcephaly_5;CLNACC=RCV000020792.1
1	197070901	A	G	CLINSIG=non-pathogenic;CLNDBN=Primary_autosomal_recessive_microcephaly_5;CLNACC=RCV000020789.1
1	197091537	A	T	CLINSIG=non-pathogenic;CLNDBN=Primary_autosomal_recessive_microcephaly_5;CLNACC=RCV000020768.1
1	197094030	C	T	CLINSIG=non-pathogenic;CLNDBN=Primary_autosomal_recessive_microcephaly_5;CLNACC=RCV000020763.1
1	203194186	C	T	CLINSIG=pathogenic;CLNDBN=Chitotriosidase_deficiency;CLNACC=RCV000010134.1
1	215901492	C	T	CLINSIG=non-pathogenic;CLNDBN=AllHighlyPenetrant;CLNACC=RCV000041709.1
1	216172380	A	G	CLINSIG=non-pathogenic;CLNDBN=AllHighlyPenetrant;CLNACC=RCV000041888.1
1	216173934	A	C	CLINSIG=non-pathogenic;CLNDBN=AllHighlyPenetrant;CLNACC=RCV000041834.1
1	2418092	T	C	CLINSIG=non-pathogenic;CLNDBN=FUT1/FUT2_POLYMORPHISM;CLNACC=RCV000000726.1
1	46870761	C	A	CLINSIG=other;CLNDBN=Drug_addiction\x2c_susceptibility_to;CLNACC=RCV000007116.2
1	47767914	G	A	CLINSIG=non-pathogenic;CLNDBN=Primary_autosomal_recessive_microcephaly_7;CLNACC=RCV000020526.1
1	66058513	A	G	CLINSIG=non-pathogenic;CLNDBN=LEPTIN_RECEPTOR_POLYMORPHISM;CLNACC=RCV000009047.1
1	98348885	G	A	CLINSIG=pathogenic;CLNDBN=Dihydropyrimidine_dehydrogenase_deficiency;CLNACC=RCV000000464.1
10	76285162	T	C	CLINSIG=pathogenic;CLNDBN=Hypermethioninemia_due_to_adenosine_kinase_deficiency;CLNACC=RCV000022444.26
12	52913830	A	G	CLINSIG=untested;CLNDBN=not_provided;CLNACC=RCV000058394.1
17	56292101	A	T	CLINSIG=probable-pathogenic;CLNDBN=Meckel_syndrome_type_1;CLNACC=RCV000059037.1
3	12645699	G	A	CLINSIG=pathogenic|pathogenic|pathogenic;CLNDBN=Noonan_syndrome_5|LEOPARD_syndrome_2|LEOPARD_syndrome;CLNACC=RCV000014985.20|RCV000014986.20|RCV000025926.1

Figure 10: ClinVar SNPs for A105A.

Chromosc	Coordinates	Ref Seq	Var Seq	Clinical Aspect
1	114377568	A	G	CLINSIG=other|other|other|other;CLNDBN=Diabetes_mellitus\x2c_insulin-dependent\x2c_susceptibility_to|Rheumatoid_arthritis|Systemic_lupus_erythematosus\x2c_susceptibility_to;CLNACC=RCV000003698.1|RCV000003699.1
1	11854476	T	G	CLINSIG=non-pathogenic|other|other|other;CLNDBN=MTHFR_deficiency\x2c_thermolabile_type|Schizophrenia\x2c_susceptibility_to;CLNACC=RCV000003698.1|RCV000003699.1
1	169519049	T	C	CLINSIG=pathogenic|untested|other|other;CLNDBN=Thrombophilia_due_to_factor_V_Leiden|Ischemic_stroke\x2c_susceptibility_to|Budd-Chiari_syndrome\x2c_susceptibility_to|Rec
1	171076966	G	A	CLINSIG=other|pathogenic;CLNDBN=Trimethylaminuria\x2c_mild|Trimethylaminuria;CLNACC=RCV000017711.1|RCV000020653.1
1	197031021	C	T	CLINSIG=other;CLNDBN=Venous_thrombosis\x2c_susceptibility_to;CLNACC=RCV000017984.1
1	197070815	T	C	CLINSIG=non-pathogenic;CLNDBN=Primary_autosomal_recessive_microcephaly_5;CLNACC=RCV000020792.1
1	197070901	A	G	CLINSIG=non-pathogenic;CLNDBN=Primary_autosomal_recessive_microcephaly_5;CLNACC=RCV000020789.1
1	197091537	A	T	CLINSIG=non-pathogenic;CLNDBN=Primary_autosomal_recessive_microcephaly_5;CLNACC=RCV000020768.1
1	203194186	C	T	CLINSIG=pathogenic;CLNDBN=Chitotriosidase_deficiency;CLNACC=RCV000010134.1
1	215901492	C	T	CLINSIG=non-pathogenic;CLNDBN=AllHighlyPenetrant;CLNACC=RCV000041709.1
1	216014049	C	T	CLINSIG=non-pathogenic;CLNDBN=AllHighlyPenetrant;CLNACC=RCV000041857.1
1	216172380	A	G	CLINSIG=non-pathogenic;CLNDBN=AllHighlyPenetrant;CLNACC=RCV000041888.1
1	216592303	T	C	CLINSIG=non-pathogenic;CLNDBN=AllHighlyPenetrant;CLNACC=RCV000041854.1
1	216595306	C	T	CLINSIG=non-pathogenic;CLNDBN=AllHighlyPenetrant;CLNACC=RCV000041833.1
1	47767914	G	A	CLINSIG=non-pathogenic;CLNDBN=Primary_autosomal_recessive_microcephaly_7;CLNACC=RCV000020526.1
1	66058441	A	G	CLINSIG=non-pathogenic;CLNDBN=LEPTIN_RECEPTOR_POLYMORPHISM;CLNACC=RCV000009048.1
1	66058513	A	G	CLINSIG=non-pathogenic;CLNDBN=LEPTIN_RECEPTOR_POLYMORPHISM;CLNACC=RCV000009047.1
1	98348885	G	A	CLINSIG=pathogenic;CLNDBN=Dihydropyrimidine_dehydrogenase_deficiency;CLNACC=RCV000000464.1
10	76285162	T	C	CLINSIG=pathogenic;CLNDBN=Hypermethioninemia_due_to_adenosine_kinase_deficiency;CLNACC=RCV000022444.26
16	28950223	C	T	CLINSIG=untested;CLNDBN=Malignant_melanoma;CLNACC=RCV000072079.2
17	56292101	A	T	CLINSIG=probable-pathogenic;CLNDBN=Meckel_syndrome_type_1;CLNACC=RCV000059037.1
2	71180250	C	T	CLINSIG=untested;CLNDBN=Malignant_melanoma;CLNACC=RCV000069000.2
3	12645699	G	A	CLINSIG=pathogenic|pathogenic|pathogenic;CLNDBN=Noonan_syndrome_5|LEOPARD_syndrome_2|LEOPARD_syndrome;CLNACC=RCV000014985.20|RCV000014986.20|RCV000025926.1
4	2833309	T	A	CLINSIG=pathogenic;CLNDBN=Fibrous_dysplasia_of_jaw,Fibrous_dysplasia_of_jaw,Fibrous_dysplasia_of_jaw;CLNACC=RCV000007985.1,RCV000007984.1,RCV000007983.1
9	71836364	G	T	CLINSIG=unknown;CLNDBN=AllHighlyPenetrant;CLNACC=RCV000037077.1

Figure 11: ClinVar matches A105B.

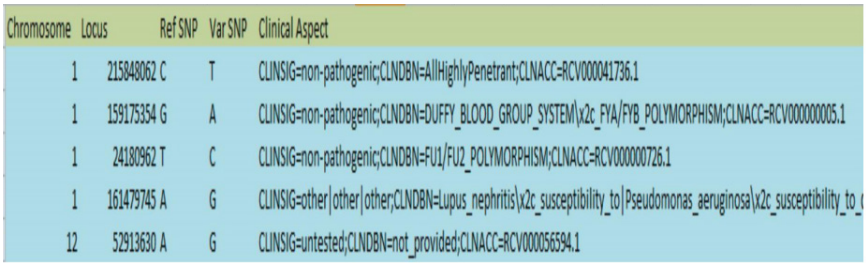


Chromosc	Coordinates	Ref Seq	Var Seq	Clinical Aspect
1	100672060	T	C	CLINSIG=pathogenic;CLNDBN=Intermediate_maple_syrup_urine_disease_type_2C;CLNACC=RCV000012727.21
1	114377568	A	G	CLINSIG=other|other|other|other;CLNDBN=Diabetes_mellitus\x2c_insulin-dependent\x2c_susceptibility_to|Rheumatoid_arthritis|Systemic_lupus_erythematosus\x2c_susceptibi
1	11854476	T	G	CLINSIG=non-pathogenic|other|other|other;CLNDBN=MTHFR_deficiency\x2c_thermolabile_type|Schizophrenia\x2c_susceptibility_to;CLNACC=RCV000003698.1|RCV000003699.1
1	154423870	A	C	CLINSIG=other|other|other;CLNDBN=Serum_level_of_interleukin-6_soluble_receptor|Serum_level_of_interleukin-6;CLNACC=RCV000015767.1|RCV000029243.1
1	161479745	A	G	CLINSIG=other|other|other;CLNDBN=Lupus_nephritis\x2c_susceptibility_to|Pseudomonas_aeruginosa\x2c_susceptibility_to_chronic_infection_by\x2c_in_cystic_fibrosis|Malaria\x2c_sev
1	169519049	T	C	CLINSIG=pathogenic|untested|other|other;CLNDBN=Thrombophilia_due_to_factor_V_Leiden|Ischemic_stroke\x2c_susceptibility_to|Budd-Chiari_syndrome\x2c_susceptibility_to|Recurr
1	171076966	G	A	CLINSIG=other|pathogenic;CLNDBN=Trimethylaminuria\x2c_mild|Trimethylaminuria;CLNACC=RCV000017711.1|RCV000020653.1
1	182554557	C	T	CLINSIG=other;CLNDBN=Prostate_cancer\x2c_susceptibility_to;CLNACC=RCV000013880.2
1	193218843	C	G	CLINSIG=non-pathogenic;CLNDBN=Parathyroid_carcinoma;CLNACC=RCV000020880.2
1	197031021	C	T	CLINSIG=other;CLNDBN=Venous_thrombosis\x2c_susceptibility_to;CLNACC=RCV000017984.1
1	197070815	T	C	CLINSIG=non-pathogenic;CLNDBN=Primary_autosomal_recessive_microcephaly_5;CLNACC=RCV000020792.1
1	197070901	A	G	CLINSIG=non-pathogenic;CLNDBN=Primary_autosomal_recessive_microcephaly_5;CLNACC=RCV000020789.1
1	197091537	A	T	CLINSIG=non-pathogenic;CLNDBN=Primary_autosomal_recessive_microcephaly_5;CLNACC=RCV000020768.1
1	197094030	C	T	CLINSIG=non-pathogenic;CLNDBN=Primary_autosomal_recessive_microcephaly_5;CLNACC=RCV000020763.1
1	203194186	C	T	CLINSIG=pathogenic;CLNDBN=Chitotriosidase_deficiency;CLNACC=RCV000010134.1
1	215848062	C	T	CLINSIG=non-pathogenic;CLNDBN=AllHighlyPenetrant;CLNACC=RCV000041726.1
1	215901492	C	T	CLINSIG=non-pathogenic;CLNDBN=AllHighlyPenetrant;CLNACC=RCV000041709.1
1	216172380	A	G	CLINSIG=non-pathogenic;CLNDBN=AllHighlyPenetrant;CLNACC=RCV000041888.1
1	216173934	A	C	CLINSIG=non-pathogenic;CLNDBN=AllHighlyPenetrant;CLNACC=RCV000041834.1
1	223285200	G	A	CLINSIG=other|pathogenic|pathogenic;CLNDBN=Legionellosis|Systemic_lupus_erythematosus\x2c_resistance_to\x2c_1|Melioidosis\x2c_resistance_to;CLNACC=RCV000007037.2|RCV000007
1	47767914	G	A	CLINSIG=non-pathogenic;CLNDBN=Primary_autosomal_recessive_microcephaly_7;CLNACC=RCV000020526.1
1	55518316	C	T	CLINSIG=non-pathogenic;CLNDBN=Familial_hypercholesterolemia;CLNACC=RCV000030351.1
1	55529237	G	A	CLINSIG=non-pathogenic;CLNDBN=Familial_hypercholesterolemia;CLNACC=RCV000030340.1
1	63872032	T	C	CLINSIG=pathogenic|non-pathogenic;CLNDBN=Congenital_disorder_of_glycosylation_type_1C|AllHighlyPenetrant;CLNACC=RCV000023375.1|RCV000023558.1
1	98348885	G	A	CLINSIG=pathogenic;CLNDBN=Dihydropyrimidine_dehydrogenase_deficiency;CLNACC=RCV000000464.1
10	76285162	T	C	CLINSIG=pathogenic;CLNDBN=Hypermethioninemia_due_to_adenosine_kinase_deficiency;CLNACC=RCV000022444.26
12	52913830	A	G	CLINSIG=untested;CLNDBN=not_provided;CLNACC=RCV000058394.1
17	28123123	G	A	CLINSIG=untested;CLNDBN=Malignant_melanoma;CLNACC=RCV000061169.2
17	56292101	A	T	CLINSIG=probable-pathogenic;CLNDBN=Meckel_syndrome_type_1;CLNACC=RCV000059037.1

Figure 12: ClinVar matches A105C.

Chromosome	Locus	Ref SNP	Var SNP	Clinical Aspect
1	215848062	C	T	CLINSIG=non-pathogenic;CLNDBN=AllHighlyPenetrant;CLNACC=RCV000041736.1
1	55529187	G	A	CLINSIG=non-pathogenic;CLNDBN=Familial_hypercholesterolemia;CLNACC=RCV000030349.1
1	182554557	C	T	CLINSIG=other;CLNDBN=Prostate_cancer\x2c_susceptibility_to;CLNACC=RCV000013880.2
1	154426970	A	C	CLINSIG=other|other;CLNDBN=Serum_level_of_interleukin-6_soluble_receptor|Serum_level_of_interleukin_6;CLNACC=RCV000015767.1|RCV000
1	161479745	A	G	CLINSIG=other|other|other;CLNDBN=Lupus_nephritis\x2c_susceptibility_to|Pseudomonas_aeruginosa\x2c_susceptibility_to_chronic_infection_t
17	26110113	G	A	CLINSIG=untested;CLNDBN=Malignant_melanoma;CLNACC=RCV000063169.2
12	52913630	A	G	CLINSIG=untested;CLNDBN=not_provided;CLNACC=RCV000056594.1

Figure 13: Novel ClinVar SNPs in A105C.



Chromosome	Locus	Ref SNP	Var SNP	Clinical Aspect
1	215848062	C	T	CLINSIG=non-pathogenic;CLNDBN=AllHighlyPenetrant;CLNACC=RCV000041736.1
1	159175354	G	A	CLINSIG=non-pathogenic;CLNDBN=DUFFY_BLOOD_GROUP_SYSTEM\x2c_FYA/FYB_POLYMORPHISM;CLNACC=RCV000000005.1
1	24180962	T	C	CLINSIG=non-pathogenic;CLNDBN=FU1/FU2_POLYMORPHISM;CLNACC=RCV000000726.1
1	161479745	A	G	CLINSIG=other|other|other;CLNDBN=Lupus_nephritis\x2c_susceptibility_to|Pseudomonas_aeruginosa\x2c_susceptibility_to_
12	52913630	A	G	CLINSIG=untested;CLNDBN=not_provided;CLNACC=RCV000056594.1

Figure 14: Novel ClinVar SNPs in A105D.

DISCUSSION

The motivation of the current work is to lay the foundation for a personalized medicine scientific age, demonstrating clearly that the nuts-and-bolts needed for personalized diagnostics is well in place for us to confidently enter the commercialization stage of the technology. Although 14x coverage is far higher than the previous studies of 1000 genome project phase1, which had lower coverage, future similar studies can be conducted with coverage of a recommendation of 50x if genome assembly approach before analytics is to be deployed. The results of mtDNA analysis seemed to not be affected by these factors as 1000s of copies of mtDNA exist in the cell anyways. The clinical results generated by the analysis of non-mitochondrial DNA cannot be trusted with complete sanctity for coverage of only 14 x, and so those medical outcomes were not discussed here. Certainly, the work should not be left to be purely automated and the results generated by the methods described here should also be validated by a team of experts. The limitation of the work also questions the need for cheaper computing facilities with more high-performance memory for the purpose, especially for genome assembly work. Techniques, such as mapping, read directly to the reference genome can also be used as an alternative, which does not require a high amount of memory, but those techniques come with their own set of problems such as missing out any novel region and their genomic loci information in the test genome, which might be missing in the reference genome.

CONCLUSION

This research article improves our understanding of human genetics, variations in the genome, and inheritance. 'Mitochondrial Adam' theory can well co-exist with 'Mitochondrial Eve' theory as proposed by Lewin et al. given that we now have evidences of paternal mtDNA inheritance, while we can continue with 'Y-chromosomal Adam' theory given that both the SVs and SNPs sum of nucleotide variance and standard deviation point out to the same findings. We also pro-posed that given that the standard deviation of the sum of nucleotides in SVs are larger than that of SNPs, SVs differences would better characterize the gender of an individual. It provides us with new scopes to fetch relevant information and opens the door for many newer technologies to be built based on the discoveries. The discoveries make us more equipped with statistical and robust, efficient and relatively less costly means to derive information such as sex determination, or immunologic response to disease, or success rate of organ transplant, or susceptibility to diseases and possible cure for them. Given that there is a larger standard deviation of the sum of the nucleotides of InDels (SVs) than sum of SNPs, this can greatly and strongly impact how personalized genomic analysis and diagnostics are being carried out, as the scientific community develops algorithms and tools to better utilize this knowledge shared by this paper.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The names of individuals participating in the research were kept

anonymous, and their consent was taken for analysis work to be published. No medical out-come of the work was reported. This paper did not need any additional ethics committee approval written or verbal, since this is an extended version of papers, and the same data was used, for more detailed interpretation of results which has been reported in this paper.

CONSENT FOR PUBLICATION

Consent to publication was taken at the time when bio-samples were being col-lected. This paper did not need any additional consent for publication written or verbal, since this is an extended version of papers, and the same data was used, for more detailed interpretation of results which has been reported in this paper.

AVAILABILITY OF DATA AND SUPPLEMENTARY MATERIALS

All supporting data are provided in this article. Supplementary materials can be downloaded from https://sites.google.com/a/iitdalumni.com/abi/educational-papers.

COMPETING INTERESTS

None to be declared.

FUNDING

This paper did not need any additional data acquisition since this is an extended version of papers, and the same data was used, for more detailed interpretation of results, and thus no additional funding was needed.

AUTHORS' CONTRIBUTIONS

All work, idea generation, coding, analysis of results and writing the paper has been done by the first author.

ACKNOWLEDGEMENTS

Author is thankful to Ms Kelin Coleman for proof-checking the manuscript before submission.

REFERENCES

1. 1000 Genomes Project Consortium. A map of human genome variation from population scale sequencing. Nature. 2010;467(7319):1061.
2. Mills RE. Mapping copy number variation by population scale sequencing. Nature Published Online. 2011;97(08):1038.
3. 1000 Genomes Project Consortium. A map of human genome variation from population scale sequencing. Nature. 2010;467(7319):1061.
4. Singh AN. The unmasking of Mitochondrial Adam and Structural Variants larger than point mutations as stronger candidates for traits, disease phenotype and sex determination. arXiv preprint arXiv. 2021; 21(02).134-69.
5. Singh AN, Unmasking of Mitochondrial Adam, Book of abstract, The 21st International Conference on Bioinformatics and Computational Biology (BIOCOMP 2020) as part of American Council on Science and Education / CSCE 2020;ISBN#1-60132-512-6.
6. Singh AN. A105 family decoded: discovery of genome-wide fingerprints for personalized genomic medicine. F1000 Research. 2012;3.
7. Abhishek Narain Singh. Customized Biomedical Informatics, BIOCOMP'13, The 14th International Conference on Bioinformatics and Computational Biology, July 22-25, 2013, Las Vegas, USA.
8. Wilson IJ, Carling PJ, Alston CL, Floros VI, Pyle A, Hudson G, et al. Mitochondrial DNA sequence characteristics modulate the size of the genetic bottleneck. Hum Mol Genet. 2016;25(5):1031-41.
9. Sallevelt SC, de Die-Smulders CE, Hendrickx AT, Hellebrekers DM, de Coo IF, Alston CL, et al. De novo mtDNA point mutations are common and have a low recurrence risk. J Med Genet.2017;54(2):73-83.
10. Li M, Rothwell R, Vermaat M, Wachsmuth M, Schröder R, Laros JF, et al. Transmission of human mtDNA heteroplasmy in the Genome of the Netherlands families: support for a variable-size bottleneck. Genome Res. 2016;26(4):417-26.
11. Singh AN. Customized biomedical informatics. Big Data Anal. 2018;3(1):1-2.
12. Singh AN. Precision Genomics & Biomedical Discoveries. In Proceedings of the International Conference on Bioinformatics & Computational Biology (BIOCOMP). The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp). 2019; pp. 74-83.
13. Luo S, Valencia CA, Zhang J, Lee NC, Slone J, Gui B, et al. Biparental inheritance of mitochondrial DNA in humans. Proceedings of the National Academy of Sciences. 2018;115(51):13039-44.
14. Schwartz M, Vissing J. Paternal inheritance of mitochondrial DNA. N Engl J Med. 2002;347(8):576-80.
15. Lewin R. The unmasking of mitochondrial Eve. Science. 1987 Oct 2;238(4823):24-6.
16. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: Public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014;42(D1):D980-5.
17. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res.2016;44(D1):D862-8.
18. Singh AN. Precision Genomics and Biomedical Discoveries. InProceedings of the International Conference on Bioinformatics and Computational Biology (BIOCOMP) The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp) 2019;74-83.
19. Fanciulli M, Norsworthy PJ, Petretto E, Dong R, Harper L, Kamesh L, et al. FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. Nat Genet. 2007;39(6):721-3.
20. Aitman TJ, Dong R, Vyse TJ, Norsworthy PJ, Johnson MD, Smith J, et al. Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature. 2006;439(7078):851-5.
21. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science. 2005;307(5714):1434-40.
22. Fellermann K, Stange DE, Schaeffeler E, Schmalzl H, Wehkamp J, Bevins CL, et al. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. Am J Hum Genet. 2006;79(3):439-48.

23. Yang Y, Chung EK, Wu YL, Savelli SL, Nagaraja HN, Zhou B, et al. Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): Low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. Am J Hum Genet.2007;80(6):1037-54.

24. Hollox EJ, Huffmeier U, Zeeuwen PL, Palla R, Lascorz J, Rodijk-Olthuis D, et al. Psoriasis is associated with increased β-defensin genomic copy number. Nat Genet. 2008;40(1):23-5.

25. Singh AN. Variations in Genome Architecture, Poster. In International Congress on Personalized Medicine 2012.

26. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet. 2006;7(2):85-97.

27. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet. 2008;40(6):695-701.

28. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nat Genet. 2004;36(9):949-51.

29. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-scale copy number polymorphism in the human genome. Science. 2004;305(5683):525-8.

30. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature. 2006;444(7118):444-54.

31. Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, et al. Fine-scale structural variation of the human genome. Nat Genet. 2005;37(7):727-32.

32. Khaja R, Zhang J, MacDonald JR, He Y, Joseph-George AM, Wei J, et al. Genome assembly comparison identifies structural variants in the human genome. Nat Genet. 2006;38(12):1413-8.

33. Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, et al. Paired-end mapping reveals extensive structural variation in the human genome. Science. 2007;318(5849):420-6.

34. Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, et al. Mapping and sequencing of structural variation from eight human genomes. Nature. 2008;453(7191):56-64.

35. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. Nature. 2010;464(7289):704-12.

36. Buchanan JA, Scherer SW. Contemplating effects of genomic structural variation. Genet Med. 2008;10(9):639-47.

37. Amberger J, Bocchini CA, Scott AF, Hamosh A. McKusick's online Mendelian inheritance in man (OMIM®). Nucleic Acids Res. 2009;37(suppl_1):D793-6.

38. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29(1):308-11.

39. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al.Pharmacogenomics knowledge for personalized medicine. Clin Pharmacol Ther. 2012;92(4):414-7.

40. Thorisson GA, Lancaster O, Free RC, Hastings RK, Sarmah P, Dash D, et al. HGVbaseG2P: A central genetic association database. Nucleic Acids Res. 2009;37(suppl_1):D797-802.

41. UniProt Consortium. Activities at the universal protein resource (UniProt). Nucleic Acids Res. 2014;42(D1):D191-8.

42. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. ABySS: A parallel assembler for short read sequence data. Genome Res. 2009;19(6):1117-23.

43. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAP de novo 2: An empirically improved memory-efficient short-read de novo assembler. Gigascience. 2012;1(1):2047-17X.

44. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2011;27(4):578-9.

45. Singh AN. Comparison of Structural Variation between build 36 Reference Genome and Celera R27c Genome using GenomeBreak, Poster Presentation. In the 2nd Symposium on Systems Genetics, Groningen 2011; pp. 29-30.

46. Singh A. Genomebreak: A versatile computational tool for genome-wide rapid investigation, exploring the human genome, a step towards personalized Genomic Medicine.

47. Helena Mangs A, Morris BJ. The human pseudoautosomal region (PAR): Origin, function and future. Curr Genomics. 2007;8(2):129-36.

48. Wiesner et al. Counting target molecules by exponential polymerase chain reaction: copy number of mitochondrial DNA in rat tissues. Biochem Biophys Res Commun.1992;183(2):553-9

49. Ekblom R, Wolf JB. A field guide to whole‐genome sequencing, assembly and annotation. Evolutionary applications. 2014;7(9):1026-42.

50. Khrapko K. Two ways to make an mtDNA bottleneck. Nat Genet. 2008 Feb;40(2):134-5.

51. Dulik MC, Zhadanov SI, Osipova LP, Askapuli A, Gau L, Gokcumen O, et al. Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. Am J Hum Genet. 2012;90(2):229-46.

52. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. Science. 2010;327(5962):167-70.