ISSN 2394 - 2053 (Print)
ISSN 2394 - 2061 (Online)
RN: 143216 (RNI)
www.ojpas.com

CONTINUING MEDICAL EDUCATION (CME) Open Journal of Psychiatry & Allied Sciences

A brief summary of human molecular genetic techniques for clinical psychiatrists

Abstract

Concerted and systematic efforts to understand genetics of human health and disease over the preceding 60 odd years have witnessed remarkable progress. The incremental gains through this journey were enabled by chromosomal analysis, recombinant deoxyribonucleic acid (DNA) techniques, notable discovery of single nucleotide polymorphisms following the Human Genome Project, consequent genome-wide variant-based studies, and now whole genome sequencing with ultimate diagnostic potential. Of note, success in prediction and prevention of chromosomal and single gene disorders comprising ~six to eight per cent each of all genetic disorders have been unprecedented but uncovering genetics of common complex disorders conferring ~60% of the genetic disease burden continues to pose a challenge and await new analytical paradigms - a mix of reductionist and organismal biology together with artificial intelligence and machine learning approaches being the current trend. A brief account of this path of progress in medical genetics and genomic insights along with limitations, to achieve the overarching goals of predictive, preventive, personalised, and participatory medicine is presented in this article.

Keywords: Gene. Genomics. Clinical. Medicine.

Chandra Bhushan Rai[1], Anirban Mukhopadhyay[2], Smita N Deshpande[3], BK Thelma[4]

[1]Department of Psychiatry, Centre of Excellence in Mental Health, Atal Bihari Vajpayee Institute of Medical Sciences & Dr. Ram Manohar Lohia Hospital, New Delhi-110001, India, [2]Department of Genetics, University of Delhi South Campus, New Delhi-110021, India, [3]Department of Psychiatry, Centre of Excellence in Mental Health, Atal Bihari Vajpayee Institute of Medical Sciences & Dr. Ram Manohar Lohia Hospital, New Delhi-110001, India, [4]Department of Genetics, University of Delhi South Campus, New Delhi-110021, India

Correspondence: Dr. Smita N. Deshpande, Professor and Senior Consultant, Department of Psychiatry, De-addiction Services & Resource Centre for Tobacco Control, Centre of Excellence in Mental Health, Atal Bihari Vajpayee Institute of Medical Sciences & Dr. Ram Manohar Lohia Hospital, Banga Bandhu Sheikh Mujib Road, New Delhi-110001, India. smitadeshp@gmail.com

Received: 12 April 2020
Revised: 7 August 2020
Accepted: 12 August 2020
Epub: 18 August 2020
DOI: 10.5958/2394-2061.2021.00010.0

INTRODUCTION

Scientists have been long engaged in uncovering the secrets in the genetic Bible of all living beings - big or small - by deciphering the language of deoxyribonucleic acid (DNA) - the core element. The last seven decades witnessed a stepwise progress in our understanding of biology in general but all centring around DNA, the molecule of life. In humans, particular interest from the angle of understanding health and disease was the focus.

Human genetic disorders can be broadly grouped into chromosomal disorders (~six to eight per cent, examples - Down syndrome, Turner syndrome, etc.), single gene (monogenic or Mendelian) disorders (also ~six to eight per cent, examples – Duchenne muscular dystrophy, cystic fibrosis, β-thalassaemia, etc.), and common complex disorders (~60%, examples - type 2 diabetes, cardiovascular diseases, hypertension, schizophrenia, rheumatoid arthritis, etc.). The tremendous advances and achievements in medical genetics and biomedical genomics contributed substantially to the understanding of the aetiology of chromosomal disorders and considerable proportion of monogenic disorders. But quantifiable progress in genetics of common complex traits has been limited and awaits new analytical paradigms.

This path of progress evolved from the big picture - from the human disorders with distinct phenotypes or manifestations down to the tiniest component - the single nucleotide in the DNA of all individuals; via a range of distinct, identifiable organisational intermediates. These include chronologically the chromosomes; chromosomal banding patterns; recombinant DNA techniques enabling placing small segments of DNA in clones; identification of unique and repetitive sequence signatures in DNA segments as markers; Human Genome Project (HGP) and discovery of

single nucleotide polymorphisms; and finally, to the current capabilities, which include next generation sequencing at a very fast pace for diagnosis (Figure 1).

A brief overview of this journey of genetics in health and disease through the last five to six decades, the contemporary tools in medical genomics, distinct eras in genomics research and genomic insights at hand, particularly for the common complex disorder category, is presented in this article.

TOOLS FOR GENE MAPPING

Different levels of organisation of the DNA sequences and a range of sequence signatures, commonly referred to as *genetic markers*, are nothing but different handles which can be used for the sole aim of identifying the minimal DNA segment which is sufficient to contribute to a specific trait or feature, and termed as a gene (biochemical markers were also used early on for gene mapping but that is beyond the scope of this article). The method to look for a gene using these markers is termed as physical mapping and is analogous to a postal address both in its structure and function, in that it helps locate a gene of interest.

If one were to document the landmarks or specify the distinct phases of this journey, they may be termed as organismal (phenotype-based/pre-chromosomal); chromosomal; recombinant DNA; HGP; genome-wide associations; next generation sequencing; and now big data and artificial intelligence. These tools and techniques help in understanding the structure and organisation of the total genetic material in an individual, termed as the genome. The same tools are used, singly or in combination, to analyse an individual affected with a disease detectable phenotypically (such as Down syndrome or intellectual disability) or at the level of symptoms (such as thalassaemia or cystic fibrosis or diabetes or cardiovascular disease or schizophrenia). However, the need is to search for the difference(s)/variations in the genetic material (at chromosomal or DNA levels) between the ill and the well.

PRE-DNA AND PRE-GENOMICS ERA

Documentation of human diseases and the pattern of the inheritance in the respective families (commonly referred to as Mendelian disorders) by family physicians was the richest resource in the years prior to the unravelling of DNA structure in 1953.[1] Common among these were a large number of families with diseases segregating in a particular pattern - either seen in every generation or noted to skip generations or affecting only males or all children affected but passed on only through the mother. Accordingly, these disorders were labelled as autosomal dominant or autosomal recessive or X-linked or mitochondrial and so on. Such extensive documentation of phenotypes in the absence of any other experimental tools was very important and useful, and characterised that period of observational biology.

CHROMOSOMAL STUDIES

Solving the structure of DNA by Francis and Crick marked a major revolution. It provided the much-needed understanding of the basic structure of DNA, the four basic building blocks (adenine, guanine, cytosine, and thymine), of the double helical DNA molecule. These molecules are organised into a chromosome which is present in the nucleus of every somatic cell (except red blood cells [RBCs] which do not have a nucleus) which constitute tissues and finally organisms (Figure 2). Unravelling the DNA structure, however, did not at that point in time help us understand or interpret the genetic basis of disease.

It was not until after 1960 when culturing of human T lymphocytes became possible (with the addition of a mitogen), in the laboratory that chromosomes could be prepared and metaphase chromosomes visualised under a microscope. With this technique, the number of chromosomes, their shape and size, numerical and structural changes, if any, became very clear. Thus, the chromosomal basis of a group of disorders started to be understood. This branch of genetics - called cytogenetics - has the distinction of being the first tool to be used for diagnostics and remains to

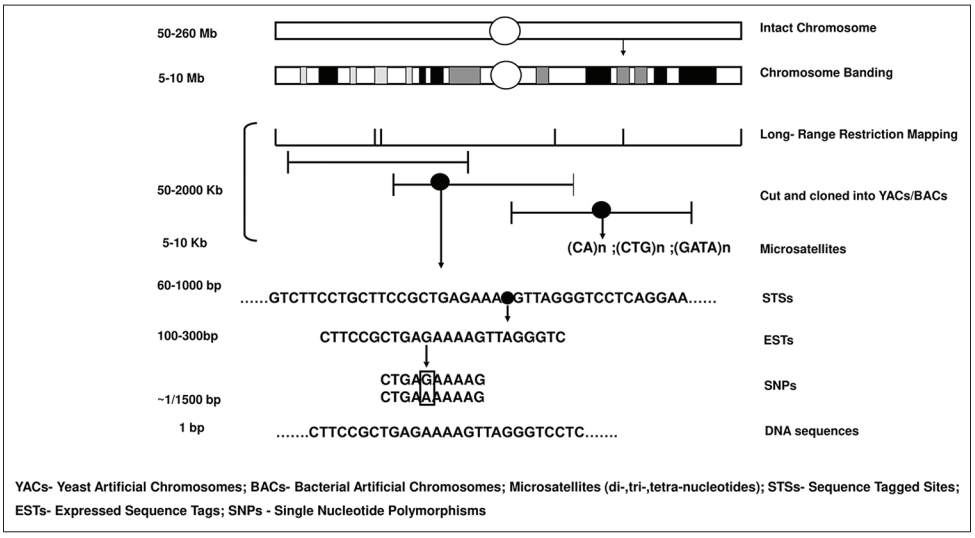

Figure 1: The commonly used genetic markers for physical mapping of the human genome with their respective degrees of resolution (a chromosome being the largest and deoxyribonucleic acid [DNA] sequence the smallest, at a single base level).

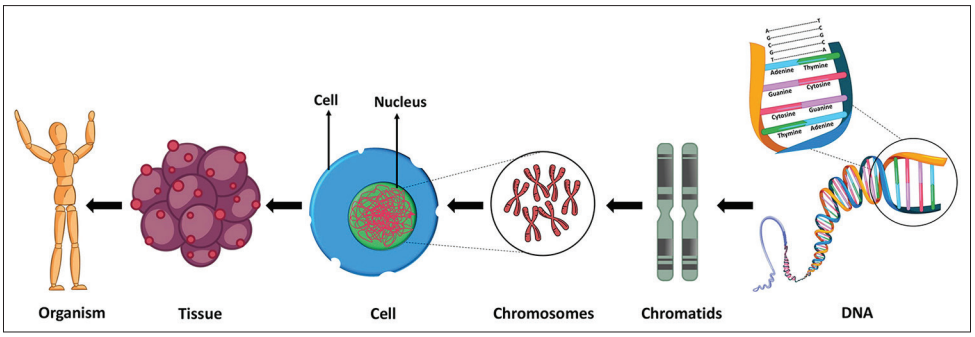

Figure 2: A cartoon depicting the relation of a double helical deoxyribonucleic acid (DNA) molecule to the ultimate human phenotype through the intermediate stages (number of chromosomes do not represent actuals but only representative).

be extensively used even today, e.g. for intellectual disability, infertility, pre-implantation diagnostics, or cancers.

With chromosomal analysis came several additional techniques such as chromosome banding which enabled reliable karyotyping of humans (and also a wide range of other organisms) which confirmed the diploid chromosome number as 23 pairs (n=46; 22 pairs of autosomes and XX or XY, the sex chromosomes) which constitute a human diploid cell. The unique chromosome banding pattern of each pair enabled unambiguous identification of each one of these pairs of chromosomes and the X and the Y chromosomes.

It also enabled unambiguous identification of any numerical or structural anomaly characterising chromosomal disorders, such as an extra chromosome 21 (trisomy) in Down syndrome or only one X chromosome in Turner syndrome (45 XO) or 47 in Klinefelter syndrome (XXY) and so on. A gross structural anomaly such as a deletion or duplication of a segment of a chromosome could also be recognised based on the comparative banding patterns between a normal and affected individual. With improvements in this technique, e.g. fluorescence *in situ* hybridisation (FISH), it remains a powerful diagnostic tool for a whole range of chromosomal disorders which constitute ~six to eight per cent of all human genetic disorders.

RECOMBINANT DNA TECHNOLOGY

Late 1960s through 1970s marked the next memorable milestone for all branches of biology in general and genetics including human genetics, in particular. This period was marked with flurry of activities with restriction enzymes, DNA cloning, amplification of genomic regions of interest (with polymerase chain reaction or PCR) from the total genome combined with DNA sequencing. This marked the beginning of molecular genetics. A researcher could take DNA of a healthy or affected individual, cut them with restriction enzymes, clone them into suitable vectors (plasmids, cosmids, etc.), perform additional experiments such as Southern blots and finally sequence clones of human DNA fragments.

With this genetic tool box, it became possible to identify several disease causal genes using large, informative families with the respective disease - the first being dystrophin gene for Duchenne muscular dystrophy (DMD), an X-linked recessive condition,[2] wherein the X chromosome carrying the mutation is generally passed on from the carrier mother to a son who would then be affected. Cystic fibrosis, an autosomal recessive condition was the next, for which the gene, cystic fibrosis transmembrane conductance regulator (CFTR) was identified.[3] This was followed by identification of fragile X mental retardation 1 (FMR 1) gene for the fragile X mental retardation, again an X-linked condition[4] and so on.

Thus, considerable advancements were witnessed with detection of disease-causing genetic changes (mutations) in various genes for several single gene disorders (monogenic disorders) but little progress was made for common complex traits which include type II diabetes, hypertension, cardiovascular diseases, schizophrenia, Parkinson's disease, etc. Some early biochemical and pharmacological evidence was available, with implications for the possible underlying pathology in these common conditions. However, there were neither genetic leads to suggest the disease causal or risk conferring genes and their likely numbers, nor definitive environmental cues to understand the aetiology of this common complex group of disorders, which constitute approximately 60% of all human genetic disorders and continue to pose a challenge.

HGP AND SINGLE NUCLEOTIDE POLYMORPHISMS

Encouraged by the success with genetics of single gene disorders, improved recombinant DNA techniques including large scale DNA sequencing feasibility, the landmark HGP with the main aim to sequence the three billion base pairs in a human haploid genome (as in sperm or egg, unlike diploid in a somatic cell which arises by the fusion of the gametes), was initiated in 1987. HGP was a joint effort of 16 different countries to be carried out over following 15 years (India was not a part).

The investigators believed that sequencing the entire human genome of a few representative individuals, would yield information on all the genes in a genome, whose functions could be studied, all disease-related genes could then be identified enabling understanding the biology of diseases. However, on completion of the project three years prior to the original schedule, complete answers were still elusive. A rough estimate of total genes (very different from earlier beliefs) was obtained but their functional relevance and association with diseases remained a distant goal. Nevertheless, the most notable fallout of HGP was the systematic documentation

of a very large number of single base changes distributed throughout the genome - the discovery of *single nucleotide polymorphisms* (SNPs).

Today we know of more than ten million commonly found SNPs on an average in an individual and they serve as powerful genome-wide physical map markers. In other words, ~99.9% of the sequences across human genomes would be similar and only ~0.1%, would show a base change which is akin to a spelling mistake which may change the amino acid (see triplet codon in Figure 3 and appendix) and therefore the protein structure and consequently its function or truncate the protein itself. Or, it may be an alternate word but broadly conveying the same meaning; or a totally different one, or it may also be another way of spelling the word, without losing its meaning (such as characterise versus characterize).

It is important to note that it is this small percent of the genome with such base changes which makes each individual unique, by contributing to differences in the sequence composition of proteins and other regulatory signatures. These variations distinguish one individual from another, two siblings from each other, and so on, except for monozygotic twins who have 100% similarity in their genomes. These single base changes also confer susceptibility to common complex diseases. With this opportunity at hand, the search was immediately on to catalogue SNPs across the human genome, across trans-ethnic populations, and across genomes of individuals affected with complex diseases and matched healthy controls. Needless to say, this offered great promise for the elusive complex trait genetics.

CANDIDATE GENE STUDIES

Based on the prior biochemical or pharmacological evidence, several *candidate genes* were selected to look for SNPs within them to be used as markers and assessed (genotyped) to establish the differences in their frequency between cases and controls. For example, in view of the documented role of dopamine in schizophrenia or Parkinson's disease, genes involved in the dopaminergic pathway such as dopamine synthesis, transport, storage metabolism, reuptake along with dopamine receptor genes became favourite candidates for testing their association with disease (Figure 4). These studies were called *candidate gene association studies* or a *hypothesis testing approach*. With this strategy, mid 90s until 2007 witnessed a large number of candidate gene-based association studies.[5]

But what emerged was not so promising because there were notable differences or non-replications in inter- and intra-population studies. For example, what was observed in a north Indian schizophrenia cohort was not replicated in the south Indian cohort, Caucasian findings were different from Indian or African population-based analyses, and so on.[6-8] These findings suggested on one hand that there could be population specific differences in the genetic variants contributing to the disease and on the other, that there could be other more important risk conferring genes yet to be identified. This obviously warranted the search for newer tools and newer paradigms in complex trait genetics.

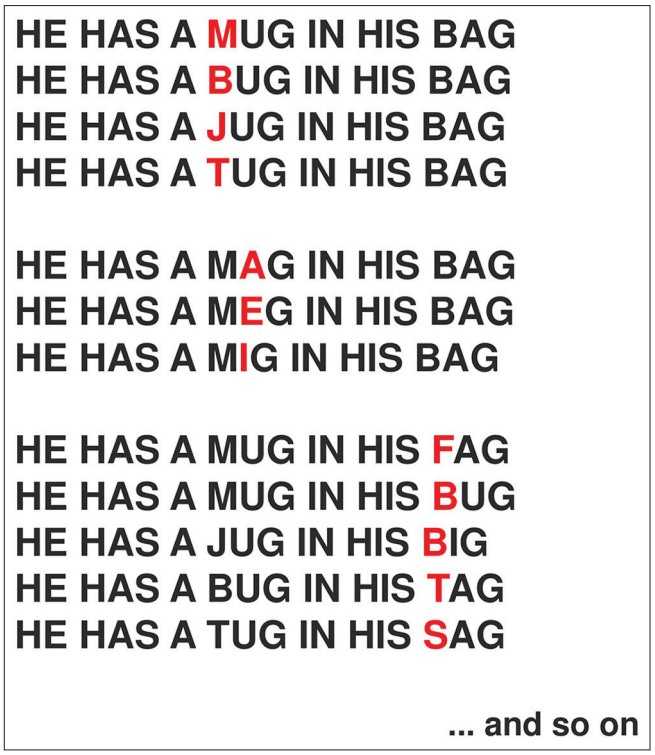

Figure 3: A cartoon to depict the role of one or more single nucleotide polymorphisms (SNPs) in a deoxyribonucleic acid (DNA) segment (It is generally a G to A or A to G; and C to G or G to C change in DNA. In rare cases, you may have any of the four bases at a given location). Lines show i) one spelling change (analogous to a SNP in a DNA segment) which can lead to no change in the meaning or nonsense change or missense change of the sentence; and ii) 2-3 changes (SNPs) which can lead to multiple missense or nonsense changes (analogous to codon changes in amino acids leading to mutant or dysfunctional or truncated proteins).

POST-GENOMIC ERA

With time more and more SNP data were generated from analysis of much larger sample sets across population groups. Then came the remarkable technological advancement of generating a chip with a large number of SNPs which could in one go, evaluate a very large number of SNPs per sample and for a large number of samples. With this, complex trait genetics took a big leap into a genome-wide search for risk genes termed as hypothesis free or hypothesis generation approach for establishing association between a marker and the disease under study.

GENOME-WIDE ASSOCIATION STUDY

Popularly referred to as GWAS, it relies on the application of chi-square statistics to analyse the genotypes of a large number of SNPs in large case-control cohorts with setting the significance value very high to account for multiple comparisons. With this revolutionary technique, new and big data began to emerge from thousands of GWASs which were (and continue to be) performed for different complex diseases in different populations. However, this approach has also yielded little consensus on the genetic players underlying the respective diseases.

To sum it up, a large amount of data for a large number of diseases from different populations, though predominantly

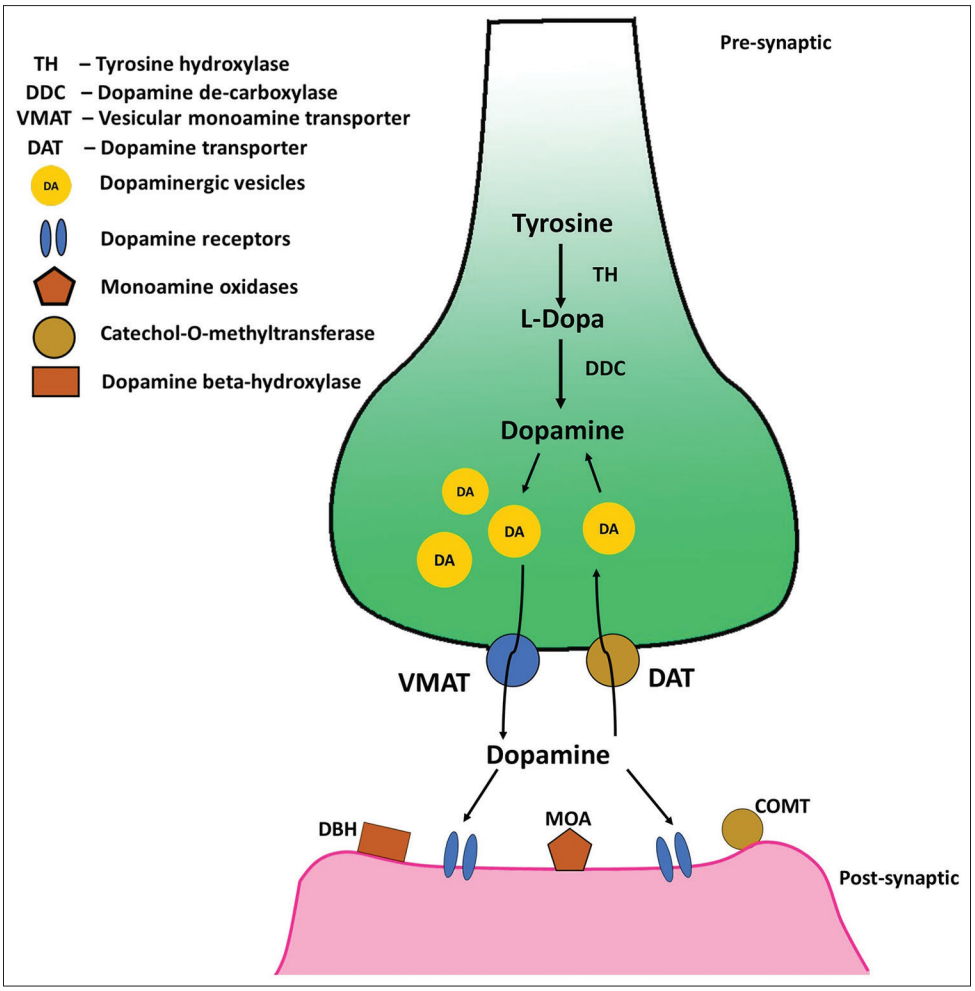

Figure 4: Candidate genes in the dopaminergic pathway. Genes code for dopamine synthesis (TH), transport, reuptake (DAT), metabolism (MAO, COMT, DBH), and receptors. MAO=monoamine oxidase, COMT=catechol-O-methyltransferase, DBH=dopamine β-hydroxylase.

Caucasian, have accumulated, but to date there is no single marker which can be used for prediction of an individual susceptibility to develop a disease, even as common as type 2 diabetes. With some insights but not sufficient for disease risk prediction and prevention, the search for alternate tools and paradigms to enable better understanding of common complex disease genetics engages genome researchers.

NEXT GENERATION SEQUENCING

Just a couple of years after the *genome-wide chip for association studies* became available - along with the realisation of its inability to deliver answers for complex traits - the next technological achievement for a sequencing approach namely the *next generation sequencing* (NGS) method emerged. This fascinating technology has now transformed the way in which we conduct human molecular genetics and biomedical genomics studies, and seems to offer promises not only for the proportion of single gene disorders hitherto not understood, but also for the elusive complex disorders. Every individual in principle can be sequenced to obtain information on all the protein coding DNA segments in a genome (*whole exome sequencing* - WES), and all of the DNA segments with non-coding/regulatory/as yet unknown functions. This can be performed for every individual who may be enrolled in the study or would just like to have his genome sequenced for medical and non-medical purposes. We have the 'ultimate' in sequencing technology in hand but with inadequate understanding of the functional significance of the largest part of the genome comprising all the non-coding/regulatory/unknown function sequences, genotype-phenotype correlations for at least complex traits seem distant.

CONCLUSION

In the light of over six decades of constant progress in developing new technologies and consequently our understanding of human genetics and diseases, where do we stand today in disease risk prediction and prevention of genetic disorders? We have routine but very effective and reliable chromosomal testing for prenatal diagnosis of several syndromic conditions and also preimplantation diagnostics in in vitro fertilisation (IVF) clinics. Identification of single base changes across the genome has been the most remarkable discovery in the last few decades and technological advances such as NGS have enabled identification of a large number of disease-causing genetic variations particularly in single gene disorders, even the rare conditions. Molecular diagnostics

for those are now very routine. But much remains to be done to understand or to predict and prevent common complex disorders. We have the possibility to perform whole genome sequencing (WGS) for individuals afflicted with a common complex trait but only to contribute to the genetic canvas of complex traits.

We now aspire for newer tools such as machine learning or artificial intelligence to make sense of the huge data thus generated. Using these methods in this next level of big data analysis, we need to analyse extensive phenotypic/clinical information, detailed genetic data and imaging, and other investigative data where relevant. Genetic aetiology of complex disorders thus continues to be a black box, warranting new world paradigms, new world analytical tools, better understanding of gene-environment interactions, etc.

Currently favourite approaches for complex disorders include the next level of family-based studies wherein multiple members affected with a disease along with the availability of a few unaffected members from the same family are taken up for NGS (evoking memories of the earliest family-based phenotypic and heritability studies). Data thereof have begun to uncover several disease-related pathways. Thus, they offer hope for personalised medicine to be put into practice in a routine clinical set up. These findings may enable repurposing of drugs based on the pathways which may have a substantial number of alterations in an affected individual or additional affected members in the family.

While these are being actively researched, our ability to identify and evaluate the contribution of genes in a few other situations such as pre-prescription testing or pharmacogenetic testing have witnessed some success and translation to the clinics. Drug response including efficacy, toxicity, etc. like the disease itself, is also very complex. But fewer genes are implicated in the pharmacokinetics and pharmacodynamics of a drug molecule and therefore, have yielded some translatable results, such as prediction of responders and non-responders, adverse drug reactions, drug dosage, etc. Another area where identification of genes and SNPs have helped has been in our understanding of the differences in the genome architecture of different populations.

To conclude, currently we have the capabilities to sequence the whole genome of every individual. What we still lack is the knowledge and ability to interpret the functional relevance of a large part of this sequence data. The functional significance of protein coding sequences is reasonably understood for less than two per cent of the total genome. All genetic errors/mutations in several of these genes which may cause disease are unknown in large number of even monogenic disorders. For the complex disorders (including schizophrenia) we still do not know (i) how many genes contribute to a disease; (ii) do qualitative (variant or mutated protein) or quantitative (differences in the amount of the protein) changes in these genes tilt the balance between health and disease; (iii) nature of gene-gene interactions; (iv) gene-environment interactions in disease aetiology; and more.

In spite of major strides in biomedical genetics in the last decade, discovery genomics continues to occupy a pivotal place in both common, and rare, human genetic disorders. This effort is complemented by computational genomics on one hand and functional genomics including cellular models of disease on the other, to achieve the goals of predictive, preventive, personalised, and participatory (P4) medicine.[9]

REFERENCES

1. Watson JD, Crick FH. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. J.D. Watson and F.H.C. Crick. Published in Nature, number 4356 April 25, 1953. Nature. 1974 Apr 26;248:765.
2. Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener C, Kunkel LM. Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell. 1987;50:509-17.
3. Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, *et al*. Identification of the cystic fibrosis gene: genetic analysis. Science. 1989;245:1073-80.
4. Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, *et al*. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell. 1991;65:905-14.
5. Pasche B, Yi N. Candidate gene association studies: successes and failures. Curr Opin Genet Dev. 2010;20:257-61.
6. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. Nat Rev Genet. 2010;11:356-66.
7. Prasad S, Bhatia T, Kukshal P, Nimgaonkar VL, Deshpande SN, Thelma BK. Attempts to replicate genetic associations with schizophrenia in a cohort from north India. NPJ Schizophr. 2017;3:28.
8. Deshpande S, Prasad S, Semwal P, Bhatia T, Nimgaonkar V, Thelma BK., Replication study of GWAS and other strongly associated markers from chromosome 6 in North Indian population. Eur Neuropsychopharmacol. 2017;27 (Suppl 3):S457.
9. Sagner M, McNeil A, Puska P, Auffray C, Price ND, Hood L, *et al*. The P4 health spectrum - a predictive, preventive, personalized and participatory continuum for promoting healthspan. Prog Cardiovasc Dis. 2017;59:506-21.

Rai CB, Mukhopadhyay A, Deshpande SN, Thelma BK. A brief summary of human molecular genetic techniques for clinical psychiatrists. Open J Psychiatry Allied Sci. 2021;12:55-61. doi: 10.5958/2394-2061.2021.00010.0. Epub 2020 Aug 18.

Source of support: Nil. **Declaration of interest:** None.

APPENDIX

A codon table with triplet codons, their redundancy, and corresponding amino acids (for ease of reference to understand single nucleotide polymorphisms [SNPs] described in Figure 3).

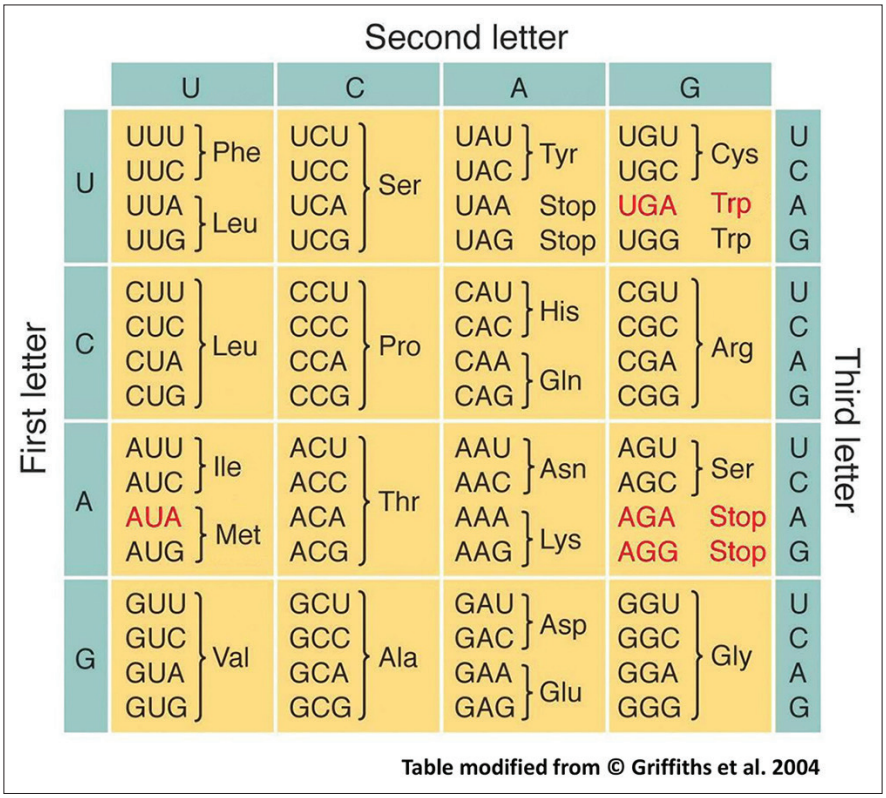

Table modified from © Griffiths et al. 2004