Bioinformatics advances Genetic Engineering Applications in Gene Therapy: A Review

Mamta sagar, Gaurav Singh, Brijendra Singh, Neetesh Pandey
Department of Bioinformatics, University Institute of Engineering and Technology,Chhatrapati Shahu Ji Maharaj University, Kanpur,
mamta10602@gmail.com

***Abstract*—Gene therapy involves repair or replacement of mutated genes. Gene expression can also be altered in order to regulate the immune system. Gene therapy requires, characterisation of defective gene, use of vector system for insertion of gene and recombination between defective and functional gene. This study reviews the application of bioinformatics and genetic engineeringmethods in gene therapy. First case of Ex vivo gene therapy is hypercholesterolemia. A mutation in LDL receptor gene on chromosome 19 is responsible for a genetic disease hypercholesterolemia which was treated by altering the expression of LDL gene by beta actin promoter from chickens and Cytomegalovirus. In vivo gene therapy is done to treat the tissue at the location of affected organ.Duchenne Muscular Distrophy and Cystic Fibrosis are treated by in vivo gene therapy where lung, pancreas and skeletal muscles are affected. Various viral, non-viral and naked DNA vector and vehicle system are used to deliver transgene into the patient body. Liposomes, Lipoplexes and polyplexes are example of non-viral vector system. Expression of defective genes may be interrupted by RNA interference. This mechanism can be used to silent the defective gene using si RNA (short interference RNA) in various disease cancer, autoimmune diseases, dominant genetic disorders, and viral infections. Role of Genomics is discussed, comparative study of genes and genomes leads to identification of hypothetical genes which may be involved in disease. In Pharmacogenomics, analysis of individual and population genetic profile is used in the treatment of Human populations. Microarray Gene expression profiling is done to analyse to study the altered gene expression in various diseases. Supervised and unsupervised clustering methods are used to analyse the microarray data.**

Keywords—Gene tharapy; Bioinformatics applications; gene therapy techniques; genetic engineering applications; Phramacogenomics; RNA interferenceC;Vector- Vehicle systems

I. INTRODUCTION

In Gene Therapy, defective gene is replaced with functional copy (transgene) of correct gene toprevent disease like AIDS, Gaucher disease, rheumatoid arthritis, a1-antitrypsin deficiency, and others. Gene therapy involves repair or replacement of mutated genes. Gene expression can also be altered in order to regulate the immune system, which leads to cell death. http://artemisinin.wizytowka.pl/. There are two approaches for gene therapy (i)Introduction of transgene into somatic cells of diseased tissue. (ii)Introduction of gene into the reproductive cells –involve ethical implicationsSomatic cells gene therapy is an acceptable practice. Introduction of gene into reproductive cells is associated with germplasm mutations that may alter the human gene. Bioinformatics play an important role in the identification of defective genes, which is facilitated using database resources and genome analysis tools. Computational study of structure and function of gens is known as genomics. These identified genes are corrected using genetic engineering approach in gene therapy.

II. GENETIC ENGINEERING IN GENE THERAPY

Gene therapy uses combined strategy to prevent the disease. This involves the following objectives.

1. Characterization of defective gene, cloning and their availability to be used in clinical study.
2. Use of vector system for insertion of gene into the correct site in the patient body. Vector is used to carry DNA sequence into the target cells by physical delivery like inhalation, injection.
3. Recombination between defective gene and inserted functional copy is necessary for expression.

Gene addition therapy is the alternative to the replacement therapy, this does not essentially require recombination between defective and inserted functional copy.

A. Ex vivo gene therapy

If diseased tissue is removed from a patient for manipulation outside the patient body, function is then restored in these cells by inserting functional gene. This therapy is suitable for blood system disease.

Normal ldl receptor gene is located on chromosome 19 human genome at 19p13.2 location, ldl gene regulates the cholesterol metabolism in blood, LDL receptors are abundantly occur at the surface of liver cells.

A mutation in LDL receptor gene on chromosome 19, raise the level of LDL (low-densitylipoprotein) which is responsible for a genetic disease hypercholesterolemia. This makes slow cholesterol metabolism. Heterozygotes have half the normal receptors; while homozygotes have from 0 to 20% of them [12].Tissues were taken outside the patient body, who was

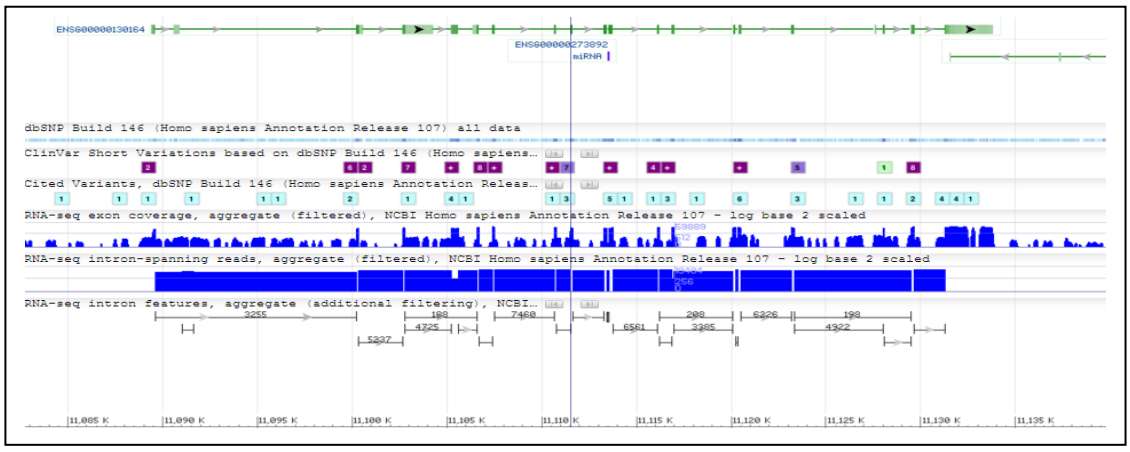

Figure 1. Mutations associated with ldr gene are depicted;annotation is retrieved from NCBI dbSNP database.

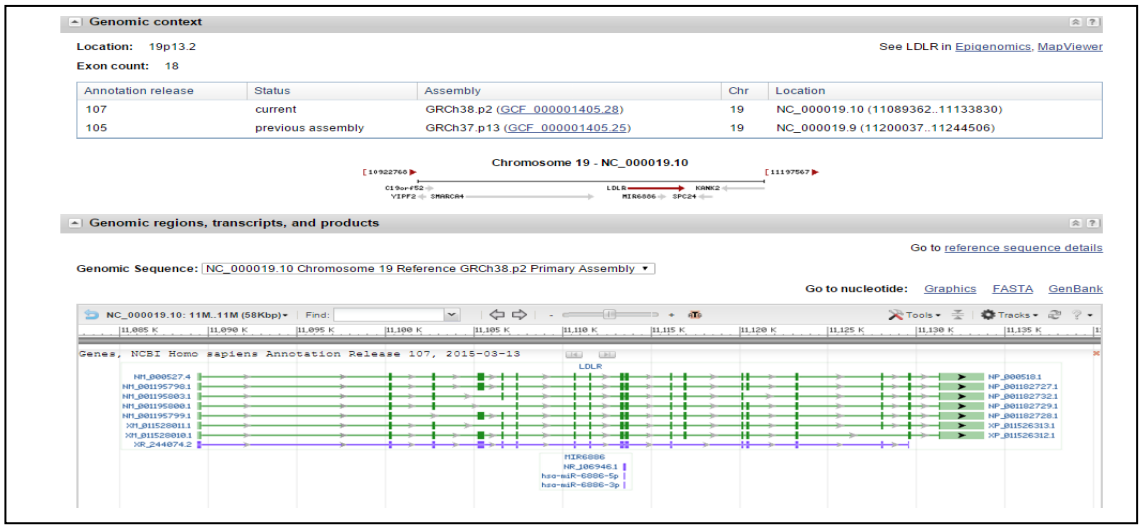

Figure2. Full annotation of ldlr gene retrieved from Genedatabase.

suffering from this disease. Expression of LDL gene was altered using beta actin promoter from chickens and Cytomegalovirus. Vector containing LDL receptor was introduced into the hepatocytes cells of Liver tissues of patient body. LDL level was decreased after introduction of modified LDL receptor gene.

Second case of application of gene therapy was Severe combined immune deficiency (immunodeficiency), a genetic disorder which is characterised by constant infections and delayed physical development [23]. A mutation in a gene on chromosome X causes Adenosine deaminase deficiency (ADA) which reduces T and B lymphocytes significantly [29].Due to conversion of deoxyadenosine into to toxic adenosine triphosphate in T cells, immunological system is severely affected [9].The gene was introduced into the LASN vector under the control of LTR(long terminal repeat) promoter. Vector alsocontains neo gene b, expressed by promoter Simian virus promoter(SMV). A recombinant breed was generated from samples of two patients and introduced into retroviral vector by transfection, then modified cells were delivered into patient body[10].Now, More than 70 mutations have been spotted which is associated with adenosine deaminase deficiency and leads to severe combined immunodeficiency (SCID). As a consequences of mutations, deficiency of the adenosine deaminase enzyme occurs, which put off normal conversion of deoxyadenosine, this ischaracterized as severe combined immunodeficiency (SCID), in ADA.

A. In vivo gene therapy

Ex vivo gene therapy is not suitable for tissue based disease (Duchenne Muscular Distrophy and Cystic Fibrosis) where lung, pancreas and skeletal muscles are affected. Treatment of disease at location of affected tissues is known as in vivo gene

Table 1. Information obtained from *Homo sapiens Annotation Release* 107, *GRCh38.p2*[35] and other references[22].

S.N.	Name of gene	Identified Mutation/ Disease	Gene Therapy	Target cells
1	ldlr (low density lipprotein receptor gene	Mutation in ldl gene two LDLR mutations leads to more severe hypercholesterolemia	Altered expression of LDL gene using beta actin promoter from CMV. Insertion of modified LDL receptor with Vector	Liver
2	Dmd(Duchenne Muscular Distrophy gene)	18 mutations in the DMD gene causes Duchenne Muscular Distrophy	Expression of dmd gene using active CMV promoter (REF) and delivery of transgene using rAAV vectorwith LTR sequence .[23]	Lungs, skeletal muscles
3	Ada (Adenosine deaminase gene)	More than 70 mutations in the ADA gene	Insertion of a gene into LASN vector (with neo gene expressed by SMV promoter) under the control of LTR promoter.	Blood
4	Cftr "cystic fibrosis transmembrane conductance regulator.	More than 1,000 mutations in the CFTR gene.common mutation delta F508	non-viral plasmid DNA and CFTR gene-liposome(25 nm) complex is used to deliver Cftr transgene	Respiratory epithelium
5	Alpha-1-antitrypsin	More than 120 mutations in the SERPINA1 gene, most common muation Glu342Lys cause alpha-1 Alpha-1-antitrypsin deficiency	Liposome	Respiratory epithelium
6	Antigen- HIV	Reduction of the Helper T-lymphocytes in the peripheral blood and opportunistic infections are associated with AIDS	Retrovirus	Blood, bone marrow
7	Genes involved in Immunity	Mutation in P53 and other tumour suppressor gene causesCancer	Retrovirus, liposome, electroporation, cellmediated transfer	Blood, marrow, tumour
8	Complement group C gene	Mutations in 15 genes can cause Fanconi's anemia, most common genes areFANCA, FANCC, and FANCG	Retrovirus	Blood, marrow
9	Glucocerebrosidase beta replacement	More than 200 mutations in the GBA gene cause Gaucher's disease	Retrovirus	Blood,marrow
10	Factor IX replacement	Mutations in the F8 (more than 1000) causes Haemophilia A and more than 900 mutaions causesHemophilia B	Retrovirus	Skin fibroblasts

11	HLA-DRB1	Variations in human leukocyte antigen (HLA) genes causesRheumatoid arthritis	Retrovirus	Synovium

therapy which involve introduction of transgene into thepatient body using vector vehicle system, gene can also be introduced directly at disease site [15].In vitro therapy require the replacement of cells and tissues from body with new gene containing cells using transfection method. Both types of therapy require different vectors to introduce new gene into cells which depend on the gene and site of disease. Deletion of one amino acid at position 508 in the CFTR protein make weak channel which can not transport chloride ions in cell membrane. This also affects water transport. As a result; cells on the surface of lungs, pancreas, and other organs create slimy mucus. This mucus hinders the air passage and glands, which is characterized as cystic fibrosis.

B. Vector- Vehicle systems for gene therapy

Vector/vehicles systems are used to deliver transgene into the cells of patient body. Biological system must be evaluated before using any physical method.

*Viral based vector:*Viruses are preferred to deliver transgene into patient body and these are known as vector. Mainly, Retroviruses are used for this purpose. Other viral systems are Adenovirues, Adeno associated viruses and Herpes Simplex Virus (HSV).Adenovirus type 2 (Ad-2) and murine leukemia virus are common vector system which are used for gene therapy [2]

Non Viral: Liposomes and Lipoplexes are example of non viral vector system. These are used for delivery of DNA. Liposomes are encapsulated DNA in lipid micelles. Lipoplexes contain charged group on lipid surface and non immunogenic in nature. Polyplexes are complex of DNA and other protein or artificial polymers (polyamino acids) because of no charge these are easily transported in blood[5]. Lipopolyplexes are composed of DNA, polymer and liposome, polymer and liposomes are used encapsulate DNA which ensure safe delivery o DNA to the nucleus.

Peptide DNA complex are also used. Plasmid DNA containing CFTR gene was complexed with a cation on the surface of liposome and delivered into the lungs tissues of patients with cystic fibrosis. This was tested in using randomised, double-blind, placebo-controlled, phase 2b trial in patients of cystic fibrosis, taken on from 18 sites in the UK[8][13].

*Naked DNA:*Naked can be introduced into the patient body in few cases where other methods could not be used for gene therapy.Viral vectors and their application in gene therapy have been discussed comprehensively in literature [22].

D. RNA Interference

Mutation greatly affect the expression of gene and produces disease, expression of these defective gene may be interrupted by RNA interference, in which short mRNA pair with the target gene and make it silence and thus expression of mutated gene is prevented, this is RNA interference, a natural phenomenon which occurs in Plant viruses as a natural defense mechanism against host gene. This mechanism can be used to silent the defective gene using siRNA (short interference RNA) in various disease cancer, autoimmune diseases, dominant genetic disorders like Huntington's diseaseand viral infections. Application of siRNAs provides the new possibilities in treatment of incurable disease [19].

Two different, RNA based and DNA based approaches are used to induce the RNA interference in target cells.

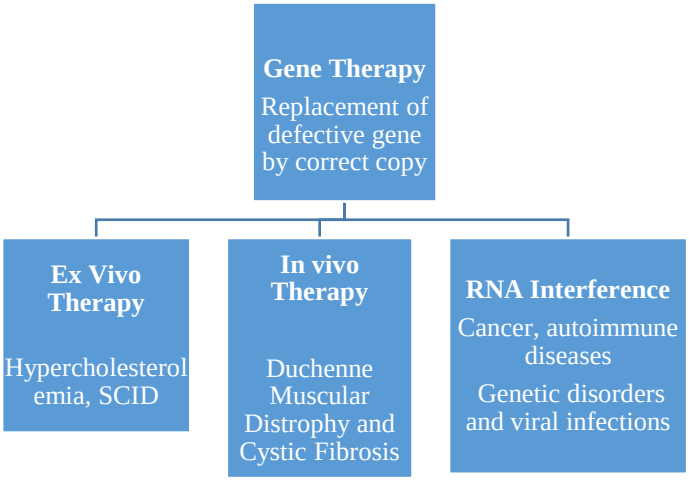

Figure 3. This chart shows classification of gene therapy approach.

In RNA based method siRNA is delivered to target cells, while in DNA based approach, a longer RNA hairpin is used to synthesise siRNA[7].

III. BIOINFORMATICS APPLICATIONS IN GENE THERAPY

The genome sequencing techniques provide the information about 20,000 to 40,000 human genes and the proteins encoded by the human genome as well as genetic variation in these genes and proteins in human populations which are involved in disease. Study of the genome of the cancer cells and tissues help us to identify new drugtargets and facilitate design of therapeutic molecule [6].Genomics approaches play an

important role in the comparative study of genes and genomes and leads to identification of hypothetical genes and protein which may be involved in disease. Analysis of individual and population genetic profile is useful in the treatment of Human populations that are sensitive for specific treatment or haveindividual response to drugs. Pharmacogenomics is study of human body response to drug based on human genetic profiles. How genetic profile affects the individual's response to drug, this study is done in Pharmacogenomics before giving any treatment to the patients. Pharmacodynamic and pharmacokinetic studies are applied in combination with Genomics to develop safe medication based on individual and population.

A. Databases and Information resources

Following database are available to help the researcher and scientist working in the related areas, which provide information required for the identification of functional and non functional gene.

1) Genetics Home References https(//ghr.nlm.nih.gov/gene):a Genetics home reference provides useful information in the

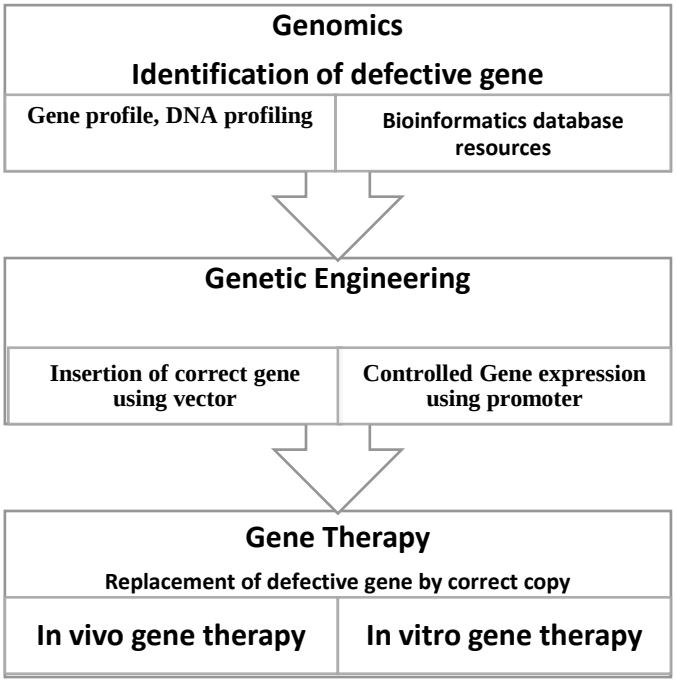

Figure 4.Flow chart illustrate the role of Genetic engineering and Bioinformatics in gene therapy

form about all the gene and product which are involved in disease.Gene orthologs, Genomic regions, transcript, SNPs, organism and Mutations

2) Gene (http//www.ncbi.nlm.nih.gov/gene/: Gene database provide information about function of gene, integrated from Reference sequences, Maps, Pathways, Variations, Phenotypes, links to Genome, phenotype.

3) OMIM (Online Mendelian Inheritance in Man)(www.omim.org):All the abnormal traits listed in this database with information about Genetic disorder, mutation, Mapping, Gene- Phenotypic, allelic variants

4) dbSNP (Single Nucleotide Polymorphism)http://www.ncbi.nlm.nih.gov/snp/: Single Nucleotide Polymorphism, insertions/deletions, microsatellites, and non-polymorphic variants.

5) UniGene: http://www.ncbi.nlm.nih.gov/unigene/: Analyses expression by tissue, age and health status and provide the information about this.

6) Geo (gene expression profile)http://www.ncbi.nlm.nih.gov/geoprofiles/: Contain information about specific gene Profile derived from microarray data. Microarray data can also be downloaded from these resources.

B. Micro Array data analysis

Gene expression profiling using microarrays is used to identify the sequence variants synthesised by splicing to study the altered gene expression in various diseases. Gene expression profile contains the expression of individual gene in different condition whereas sample profile contains the expression of the entire gene in a particular environmental condition. Supervised and unsupervised methods are used to analyse the microarray data. In unsupervised learning, similar patterns are searched in gene expression profile while supervised method uses the previous data to classify the genes.Unsupervised method include Hierarchical clustering principal component analysis[28], self-organizing maps[16], multidimensional scaling [14],[20], and singular value decomposition[32].

Supervised learning use predefined class and build the model, first training set is prepared using predefined class to predict gene classifier, this classifier is then used to classify the test data[21]

The prototype supervised method, linear discriminant analysis are commonly used to classify gene expression data (48). Neighborhood analysis and weightedvoting [30,17] are used to classify of tissue samples profile. Support vector machines methods transform data [31,33], to predict specific classes in the tissues of ovarian cancer. Research shows [26,27,4]that K-nearest neighborhood prediction can be used o classify differential expression of cancer into multiple classes.

Trained neural network models shows better predictions tumor identify and diagnose cancerous and non cancerous tumours [11, 18], and decision trees is used to discriminate among tissues of colon cancer, tree can be explored to discriminate among all the sample profile [34] and can be pruned back to limit the growth of tree using stopping rules. Genetic algorithms identify the possible mutations in the set of gene and can relate to phenotypes which may be associated with

particular disease [3]**.** The small features set algorithm uses patterns of few genes to differentiate the classes [24].
DNA profiling method and genomics approaches are discussed [1].

CONCLUSION

Ex vivo gene therapy require removal of diseased tissues from patient body and insertion of correct gene outside the patient body. A genetic disease hypercholesterolemia which was treated by altering the expression of LDL gene by Ex vivo gene therapy. In vivo gene therapy allows the treatment of tissue at the location of affected organ. Duchenne Muscular Distrophy and Cystic Fibrosis are treated by in vivo gene therapy where lung, pancreas and skeletal muscles are affected which can not be taken out from the patient body. Gene therapy has been successfully exploited to treat incurable disease using various viral, non-viral and naked DNA vector and vehicle systems to deliver transgene.
The role of Bioinformatics is important in identification of genesRNA interference is used to silent the defective gene using si RNA (shortInterference RNA) in various disease cancer, autoimmune diseases, dominant genetic disorders, and viral infections.

AKNOWLEDGMENT

This is study is supported by Department of Bioinformatics, University Institute of Information Technology, C.S.J.M. University, Kanpur. Protein Engineering and Proteomics is the part of M.Sc. Bioinformatics curriculum,which inspire us to write this article.

REFERENCES

[1] A Pfeifer, IM Verma, "Gene therapy: promises and problems. Annual review of genomics and human genetics," 2001; 2:177–211.

[2] C Fausel, "Targeted chronic myeloid leukemia therapy: seeking a cure" J Manag Care Pharm, 2007, 8, Suppl A,8–12.

[3] CH Ooi, P Tan, "Genetic algorithms applied to multi-class prediction for the analysis of gene expression data," Bioinformatics 2003;19:37–44.

[4] CL Nutt, DR Mani, RA Betensky, et al, "Gene expression-based classification of malignant gliomas correlates better with survival than histological classification," Cancer Res 2003;63:1602–7.

[5] D Scundara Manickam, D Oupicky, "Polyplex gene delivery modulated by redox potential gradients," J Drug Target, 2006, 14, 519–526.

[6] David W. Mount and Ritu PandeyMol Cancer Ther 2005;4(10). October 2005,1636-1643

[7] DH Kim, JJ Rossi, "RNAi mechanisms and applications. BioTechniques," 2008, 44, 613–616.

[8] F W Alton et al, "Repeated nebulisation of non-viral CFTR gene therapy inpatients with cystic fibrosis: a randomised, double-blind, placebo-controlled, phase 2b trial ,www.thelancet.com/respiratory" Vol 3 September 2015

[9] FJ Kim, JL Battini, N Manel, M Sitbon: Emergence of vertebrate retroviruses and envelope capture. Virology,2004, 318, 183–191.

[10] Gaspar HB, Bjorkegren E, Parsley K, Gilmour KC, KingD, Sinclair J, Zhang F et al.: Successful reconstitution ofimmunity in ADA-SCID by stem cell gene therapy following cessation of PEG-ADA and use of mild preconditioning. MolTher, 2006, 14, 505–513.

[11] J Khan,JS Wei, M Ringner, et al, "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks," Nat Med 2001;7:673–9.

[12] JE Bailey, "Toward a science of metabolic engineering," Science, 1991, 252, 1668–1675.

[13] LA Davies, , GA Nunez-Alonso, G McLachlan, SC Hyde,DR Gill,. "Aerosol delivery of DNA/liposomes to the lung for cystic fibrosis gene therapy," Hum Gene Ther Clin Dev. 2014; 25: 97–107

[14] M Bittner, P Meltzer,Y Chen et al, "Molecular classification of cutaneous malignant melanoma by gene expression profiling," Nature, 2000, 406:536–40.

[15] M Neu,O Germershaus,S Mao, KH Voigt, M Behe,T Kissel, "Cross linked nano carriers based upon poly (ethyleneimine) for systemic plasmid delivery: In vitro characterization and in vivo studies in mice," J Control Release, 2007, 118, 370–380.

[16] M Sultan, DA Wigle, CA Cumbaa, et al., "Binary tree-structured vector quantization approach to clustering and visualizing microarray data.Bioinformatics," 2002;18 Suppl 1:S111–9.

[17] MA Shipp, KN Ross, P Tamayo, et al, "Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning," Nat Med 2002;8:68–74.

[18] MC O'Neill, L Song, "Neural network analysis of lymphoma microarray data: prognosis and diagnosis near-perfect" BMC Bioinformatics 2003;4:13.

[19] N Schutze, "si RNA mechanisms. Mol Cell Endocrinol," 2004, 213, 115–119.

[20] PS Mischel, R Shai, T Shi, et al, "Identification of molecular subtypes of glioblastoma by gene expression profiling," Oncogene 2003;22:2361–73.

[21] R Simon, MD Radmacher, K Dobbin, LM McShane, "Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification," J Natl Cancer Inst 2003;95:14–8. 49. 50.

[22] Sanjukta Misra*Human Gene Therapy, "A Brief Overview of theGenetic Revolution," JAPI,february 2013 • VOL. 61,pp 127-133.

[23] S Brunner , T Sauer ,S Carotta ,M Cotten,M Saltik, E Wagner, " Cell cycle dependence of gene transfer by lipoplex, polyplex and recombinant adenovirus," Gene Ther, 2000, 7, 401–407**.**

[24] S Kim, ER Dougherty,I Shmulevich, et al, "Identification of combination gene sets for glioma classification," Mol Cancer Ther 2002;1:1229–36.

[25] S Kim,ER Dougherty,J Barrera ,Y Chen,ML Bittner, Trent JM. Strong feature sets from small samples. J Comput Biol 2002;9:127–46.

[26] S Ramaswamy, P Tamayo, R Rifkin, et al. Multiclass cancer diagnosis using tumor gene expression signatures". Proc Natl Acad Sci U S A 2001;98:15149–54.

[27] SL Pomeroy, P Tamayo, M Gaasenbeek, et al, "Prediction of central nervous system embryonal tumour outcome based on gene expression," Nature 2002;415:436–42.

[28] T Hastie, R Tibshirani, MB Eisen, et al, "Gene shaving' as a method for identifying distinct sets of genes with similar expression patterns" Genome Biol 2000;1:Research 0003, Epub 2000 Aug 4.

[29] Thrasher AJ, Gaspar HB, Baum C, ModlichU, Schambach A, Candotti F, Otsu M et al.: Gene therapy:X-SCID transgene leukaemogenicity. Nature, 2006, 443,E5–E7.

[30] TR Golub, DK Slonim, P Tamayo, et al, "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring,". Science 1999;286:531–7.

[31] TS Furey, N Cristianini, N Duffy, DW Bednarski, M Schummer, D Haussler. Support vector machine classification and validation of cancer tissue samples using microarray expression data," Bioinformatics 2000;16: 906–14.

[32] Y Kluger, R Basri, JT Chang, M Gerstein, "Spectral biclustering of microarray data: coclustering genes and conditions," Genome Res 2003;13:703–16.

[33] Yeang CH, Ramaswamy S, Tamayo P, et al. Molecular classification of multiple tumor types. Bioinformatics 2001;17 Suppl 1:S316–22.

[34] Zhang H, Yu CY, Singer B, Xiong M. Recursive partitioning for tumor classification with gene expression microarray data. Proc Natl Acad Sci U S A 2001;98:6730–5.

[35] https://ghr.nlm.nih.gov