ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY2481

Diseases Caused by Faulty Gene Expression. There are Several Hereditary Diseases that may be Entirely Cured if Detected and Treated Early Enough

Dai Yongxia[1] ; Nurul Azmir Bin Amir Hashim[2] ; Suriyakala[3]

Abstract:- Many genes and environmental factors often work together to cause a disease. The identification of significant genetic factors is useful for both medical (by aiding drug development and personalised therapy) and scientific (by shedding light on mechanistic and evolutionary aspects of illness) purposes. Linkage analysis (which joins loci that have a propensity to be inherited together) and association studies are two of the many genetic methods that have shown correlations between illnesses and particular sections of the genome (mapping correlation between alleles at different loci). Several hundreds of genes are examined in these types of studies, much too numerous to be tested experimentally as potential disease genes. The use of computer methods to assess the possibility of individual genes within a certain chromosomal area being disease genes is thus quite useful. Many diseases' susceptibilities have been demonstrated to a fair degree. changes in the rate at which genes are expressed in various cell types. In instance, if a gene or gene cluster is more common in sick individuals than in healthy people, it's likely that the gene plays a role in illness. Microarray studies were the major method for detecting the differences in expression levels.

Keywords:- *Gene Expression, Hereditary Diseases, Genetic, Disease.*

I. INTRODUCTION

When most people think of diseases with genetic origins, they typically imagine rare, single-gene disorders like cystic fibrosis (CF), phenylketonuria, or haemophilia (for example, inherited predisposition to breast cancer). There may be hundreds of rare diseases, but around 80% of them have their roots in a person's genes. One in every seventeen persons suffers from a rare illness due to the epidemic of rare illnesses. All disease processes, including the most common ones, are influenced to various degrees by the huge variances in their DNA. One or more of these factors may raise a person's risk of developing a certain disease (for example, a specific kind of cancer), while simultaneously decreasing their risk of developing a different, unrelated illness (for example, diabetes). Many diseases have environmental causes, such as poor diet and lack of exercise in the case of diabetes, but the ways in which each person's cells and body respond to these variables may differ based on their unique genetic makeup. Their susceptibility to pathogen infection is governed by their genes, and there is a wide range of responses to infection. In addition, environmental factors may have a role in the development of most malignancies since these diseases result from cumulative genetic abnormalities over the course of a lifetime. Knowledge of genetics and the whole human genome, as well as its variation in the general population, is necessary for a complete understanding of disease processes and paves the way for therapeutic interventions, preventative measures, and beneficial therapies **(Jackson, 2018).**

The development of many diseases is dependent on interactions between several genes and various environmental factors. The identification of the relevant genetic factors is useful for both medical (by simplifying drug development and personalised treatments) and scientific (by revealing mechanistic and evolutionary aspects of illness) purposes. Linkage analysis (which joins loci that have a propensity to be inherited together) and association studies are two of the many genetic methods that have shown correlations between illness and particular sections of the genome (mapping correlation between alleles at different loci). Several of the chromosomal areas studied include hundreds of genes, much too numerous to be tested experimentally as potential disease genes. Because of this, it is very useful to utilise computer methods to assess the probabilities of individual genes within a certain chromosomal area being causal for a given illness. Many disease susceptibilities have been demonstrated to a fair degree. Changes in the amount of time genes are actively being expressed in certain cell types. In instance, it's likely that a gene is implicated in illness if it's part of a group of genes that are more common in sick individuals than in healthy ones. Microarray studies were the predominant method for identifying differences in expression levels. Proteins encoded by different genes for the same disease have been demonstrated to interact with one another in certain research. It is one of the hallmarks of a disease gene that its protein product is highly correlated with the products of other disease-causing genes. A small number of prior computational techniques have used this as a jumping-off point to develop methods for identifying disease-causing genes in protein-protein interactions. Recently, there have been a number of efforts to bring together these disparate pieces of information, such as the identification of genes with variable expression and their proximity to known disease-causing genes. Protein products of disease genes are

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY2481

hypothesised to be spatially close to one another in the protein interaction network, which falls under this group of methods. This is because large-scale protein networks can only be analysed by means of approximation, greedy algorithms due to the presence of genes with fluctuating expression levels. The main difference between their method and the reported one is that theirs combines the same information without assuming that genes associated with a disease tend to cluster near genes whose expression levels have been shown to be abnormal **(Zhao, 2015).**

II. BACKGROUND OF THE STUDY:

Transcriptional regulation is an extremely crucial mechanism for controlling gene expression. The complex machinery required to exert this control is only just starting to be revealed through functional and evolutionary investigations of genomic architecture. Gene expression requires a multitude of regulatory components beyond just the promoter to occur at the proper times and in the right levels. In addition to locations upstream and downstream of the transcription unit, introns themselves may include enhancer and repressor elements. For genes with very diverse expression patterns, such as key developmental control genes, the cis-regulatory domain may extend well beyond the transcription unit. Some of the first hints were chromosomal breaks linked to disease that were discovered in seemingly unrelated parts of the genome. The great degree of conservation seen in many noncoding regions has been shown by large-scale comparisons of genomic sequences. Several of these conserved areas are transcriptional regulatory elements, recent functional investigations have indicated. Sometimes, these elements are found within distantly related neighbouring genes. Binding sites for tissue-specific, DNA-binding proteins are often found within these conserved regions. Developmental changes in chromatin conformation may regulate transcription by affecting protein accessibility to these sites. The disruption of these subtly interconnected systems has the potential to cause disease. Different from what would be expected from a change in the coding sections, the symptoms associated with a mutation in the regulatory elements would be more subtle **(Kleinjan, 2016).**

III. PROBLEM STATEMENT:

"Genetic variation may contribute to disease largely through misregulation of gene expression. Mutations in the transcription factors that control cell state may impact the autoregulatory loops that are at the core of cellular regulatory circuitry, leading to the loss of a normal healthy cell state."

This study by Lee examined a better understanding of gene expression programmes, their regulation, and the role of gene misregulation in illness need better annotation of the human genome. At the outset, it is preferable to determine whether protein-coding and noncoding genes are being actively transcribed in distinct cell types. Determining all of the genes that are expressed in one mammalian cell type presents significant difficulties. Obtaining a homogenous population of cells from most primary cell types has proven difficult, and so huge numbers of cells have been needed for such characterisation in the past. It is difficult to develop a thorough and correct annotation of ncRNA genes because of restrictions in the read length of commonly used sequencing technologies and the brief lifespan of many ncRNAs, whereas protein-coding genes are easily identified because of the existence of a coding sequence. However, new research has uncovered a great number and diversity of ncRNAs in human cells, suggesting that enhanced human genome annotation is on the horizon **(Lee, 2013).**

IV. RESEARCH OBJECTIVE:

- To find out the diseases that gene therapy can cure successfully.
- To recognize reliable is genetic testing in predicting diseases.
- To find out the four types of genetic testing.
- To explain people about genetic testing.
- To examine diseases that can be detected through genetic testing.

V. LITERATURE REVIEW:

They integrated global gene expression data from microarrays with a protein-protein interaction network the size of the whole human genome to find a strategy for prioritising disease-related genes. Since they saw that illness genes tend to cluster near other disease genes in the protein network, they suggested a Katz centrality score to account for this. Just a two-factor calibration is required to acquire the score. The ideal values for these variables may provide insight into problems with significant biological implications. The first parameter, w, modifies how the protein interaction network values differences in expression level and proximity. The possibility that a node that does not show differential expression is a disease gene is described by the second parameter, g. This provides support for the hypothesis that prioritising illness genes might be aided by data gathered from the protein-interaction network and differential expression. In contrast to the microarray data, the interaction provides extra information that may be used to predict undiscovered disease genes in the context of the study. Moreover, they optimised their approach by making advantage of limited information on established disease genes. When researchers rated all genes on a global scale rather than concentrating on individual gene loci, they discovered genes with high pleiotropy that participate in the physiological pathogenic processes of a wide variety of disorders. An further piece of evidence indicating the phenotypic interdependence, cooccurrence, and common pathophysiology of numerous illnesses is the finding of shared genes implicated in several diseases in a network environment. A new, straightforward strategy for ranking potential illness genes has been introduced in this work. Pathological phenotypes with similar genetic causes may be compared using this method **(Izum, 2016).**

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY2481

They began with the assumption that the existing literature would contradict the microarray findings. While microarray data are not affected by publication bias, it is evident that the literature has a focus that is at least influenced by the research that has come before. Their objective was to quantify this bias by comparing their conclusions to the "ground truth" given by microarray data. In retrospect, they realise how naive they were to assume a one-to-one relationship between the literature and microarray results. It is possible that the FC threshold's association with biological activity varies from gene to gene. Also, the outcomes of an expression research may change based on the FC threshold that was used. FC criteria that are too stringent, they discovered, cause researchers to use microarray expression data that may not adequately represent biological processes **(Esteban, 2017).**

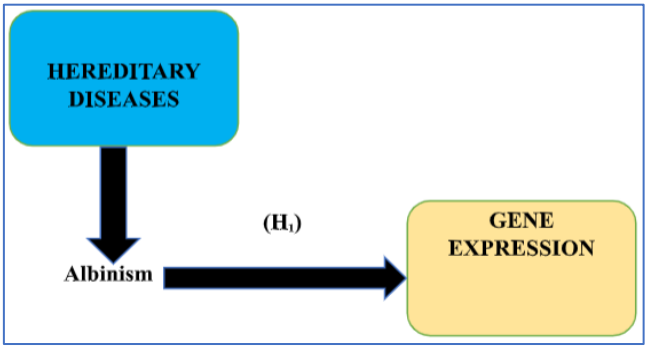

Fig 1 Conceptual Framework

VI. METHODOLOGY

➢ *Sampling:*

The subjects in this study were 600 patients sampled from the total population of the Gene Expression.

➢ *Data and Measurement:*

The data were collected during the first half of the annual year 2022. Gene expression were required. Questionnaire was distributed and quantitative analysis was implemented.

➢ *Statistical Software:*

MS-Excel and SPSS 24 Was used for Statistical analysis.

➢ *Statistical Tools:*

Descriptive analysis Was applied to understand the basic nature of the data. Validity and reliability of the data Was tested through Cronbach alpha and ANOVA.

VII. RESULT

➢ Factor Analysis

A collection of measurement items' latent component structure is often confirmed using factor analysis (FA). The observed (or measured) scores are thought to be explained by latent (or invisible) elements. The core of accuracy analysis is modelling (FA). It focuses on simulating how observed phenomena, unidentified causes, and measurement error interact. If they want to know whether the data may be utilised for factor analysis, perform the Kaiser-Meyer-Olkin (KMO) Test. To guarantee adequate sampling, both the individual model variables and the whole model are checked. The degree to which many variables may share some variance is revealed through data analysis. A smaller percentage often means that the data may be factored more easily. KMO gives values ranging from 0 to 1. Only when the KMO value is between 0.8 and 1.0 is the sample size considered to be appropriate. A KMO of less than 0.6 indicates insufficient sample and requires correction. For this reason, some writers utilise the number 0.5; between the range of 0.5 and 0.6, they must apply their best judgement.

- KMO If it's almost zero, it indicates that the total correlations are little in comparison to the magnitude of the partial correlations. I should reiterate that large-scale correlations pose a serious challenge to component analysis. Kaiser's minimal and maximum requirements are as follows: The following are Kaiser's minimal and maximum requirements. varying from 0.050 to 0.059.

Below-average (0.60-0.69) (0.60-0.69) generally in the middle school level, has a quality point value of 0.80 to 0.89. Between 0.90 and 1.00, there is amazing variation.

Table 1 KMO and Bartlett's Test:

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.973
Bartlett's Test of Sphericity	Approx. Chi-Square	3247.987
	df	190
	Sig.	.000

Finding out whether or not the data can be utilised for factor analysis is the first step in exploratory factor analysis (EFA). Kaiser suggested that the KMO (Kaiser-Meyer-Olkin) measure of sample adequacy coefficient value should be greater than 0.5 as a fundamental need for doing factor analysis in this regard. This is due to KMO, which stands for the Kaiser-Meyer-Olkin sample adequacy measure. The KMO value from this study was .973, or the data that was used. The significance threshold was also confirmed by Bartlett's test of sphericity to be 0.00.

➢ *Test for Hypothesis*

"Posing a hypothesis" is a word used in scientific discourse to describe the process of putting out a guess or assumption for the purpose of further discussion and, eventually, testing to ascertain how probable it is that the guess or assumption is true. The next step in the scientific

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY2481

process is to carry out a literature review after creating a working hypothesis. The hypothesis's prediction of the outcomes was confirmed by the outcomes. When it offers a potential solution to the inquiry's central issue, it is referred to as a hypothesis. It could be necessary to establish a lot of hypotheses, each of which would be tested, depending on the study's scope.

➢ *Dependent Variable:*

- *Gene Expression:*

The instructions in their DNA are translated into a protein via a process called gene expression. Listed below are a few instances that illustrate the significance of gene expression: Inhibition of insulin secretion as a signal for blood sugar control. Mammals' female reproductive systems inactivate their X chromosomes to avoid a "overdose" of the genes they carry. The levels of cyclin expression regulate each stage of the eukaryotic cell cycle.

✓ *Factor:*

- *Albinism:*

Hereditary diseases that are characterised by the presence of very little or no the pigment melanin in the skin, hair, and/or eyes. Albinism may cause individuals to have vision issues, hair that is white or yellow, eyes that are reddish, violet, blue, or brown, and skin that is pale.

➢ *Independent Variable:*

- *Hereditary Diseases:*

A condition that is inherited from one generation to the next and is brought about by alterations (mutations) in certain genes or chromosomes. Genetic syndromes may be passed down from either or both of a person's parents. Moreover, many members of the same immediate family (such as a mother, daughter, and sister) may be affected by the same condition.

- *The relationship between Albinism and Gene Expression:*

Autosomal recessive inheritance describes the mode of transmission of oculocutaneous albinism. The parents of a person with an autosomal recessive disorder typically each contain one copy of the mutated gene but do not exhibit symptoms of the illness.

On basis of the above discussion, the researcher formulated the following hypothesis, which was analysed the relationship between albinism and gene expression.

✓ *H_{01}: "There is no significant relationship between albinism and gene expression."*
✓ *H_1: "There is a significant relationship between albinism and gene expression."*

Table 2 Correlations

		Sum	Hl_Mean
Pearson Correlation	Sum	1.000	.995
	Hl_Mean	.995	1.000
Sig. (1-tailed)	Sum	.	.000
	Hl_Mean	.000	.
N	Sum	100	100
	Hl_Mean	100	100

Several output tables were created by doing a multiple regression analysis in SPSS Statistics. This section only describes the three essential tables that are necessary to comprehend the results of the multiple regression approach that was used to analyse their data, assuming that none of the assumptions were broken. Data from their business was exploited in this manner. Understanding the conclusion is crucial for analysing their data for the eight assumptions needed to do multiple regression, and this research, which is covered in their larger lesson, gives a detailed explanation of how to proceed. Many requirements must be met before the multiple regression procedure can start.

The table that deserves attention first is the Model Summary table. They may refer to this table, which contains the R, R2, modified R2, and standard error of the estimate, to assess the precision of a regression model.

Table 3 Model Summary

Model Summary[b]					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	1.000[a]	1.000	1.000	.000	.625
a. Predictors: (Constant), Hl_Mean,					
b. Dependent Variable: Sum					

The multiple correlation coefficient is represented by the number in the "R" column. R may be used to evaluate how well the dependent variable—disruptive innovations in this case—is anticipated. Hence, a score of 1.0 indicates a degree of prediction that is adequate. The "coefficient of determination," often known as the R2 value, is shown in the "R Square" column. This graph, which is used to infer causal relationships, displays the percentage of overall

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY2481

variance in the dependent variable that can be attributed to the impact of the independent variables (technically, it is the proportion of variation accounted for by the regression model above and beyond the mean model). The introduction of disruptive technologies is their dependent variable, and because their value is 1.0, it may be considered that their independent variables sufficiently account for the variation in this variable. Yet in order to present their findings in a manner that is professional, they must have a solid understanding of the "Adjusted R Square" (adj. R2). Researchers examine the circumstances that lead to these discoveries as well as the outcomes in an enhanced multiple regression course.

Table 4 Anova

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	55705.310	4	13926.327	104966738164406 74.000	.000[b]
	Residual	.000	95	.000		
	Total	55705.310	99			
a. Dependent Variable: Sum						
b. Predictors: (Constant), H1_Mean,						

The value for the multiple correlation coefficient is shown in the "R" column (R). R may be used to assess the dependent variable's prediction power, in this example, disruptive innovations. This example shows that a prediction accuracy of 1.0 is acceptable. The "R Square" column of the analysis of variance (ANOVA) table displays the F-ratio (R2). If this number is high, the regression model as a whole closely approximates the data. The table demonstrates that the independent variables and dependant variable have a very significant predictive association (F (5, 94) = 10496673816440674, p.0005). (Or, to put it another way, the regression model correctly explains the data.)

The basic equation that may be used to anticipate disruptive technology based on Albinism, Angelman syndrome, Apert syndrome, Cystic fibrosis: The likelihood of including essential components, Gene Expression= 1.677+ (9.343E-7 x H1_Mean (Albinism))

VIII. CONCLUSION

Several categorization methods have been developed in an attempt to define CRC cancer based on gene signatures. CMS1 represents MSI immune activation, CMS2 canonical WNT and MYC activation, CMS3 metabolic dysregulation, and CMS4 EMT high and immunological inflammation as distinct molecular subtypes of CRC. This paradigm for identifying specific tumour types allows for more targeted therapy interventions. CRIS, another categorization approach, helps quantify intrinsic signs of cancer by normalising stromal heterogeneity. As compared to CMS, the CRIS signature was found to be more reliable since it allowed for more precise geographical and temporal categorization. Retrospective analyses employing bioinformatics, FFPE, and frozen tissues have led to the discovery of gene expression scores with prognostic and predictive potential in a number of investigations during the last decade. While these approaches have evolved, nevertheless, there is still a need to find new prognostic indicators with better accuracy and prognostic potential that may encompass presently non-responding group or unclassified individuals. For instance, new prognostic biomarkers are needed to classify stage II and stage III patients, which would help in identifying those who might benefit from adjuvant treatment. Prognostic indicators may aid in the success of immunotherapies and open up a window for innovative treatments at later stages. While the identification of important biomarkers has been substantially sped up because to the availability of several bioinformatics platforms, there is an immediate need for clinical validation of these signatures prior to their application in clinical settings. The prognostic scores of many studies have been verified in external data sets, lending credibility to these biomarkers but also calling for confirmation of their validity in larger cohorts or prospective research. In addition to TNM staging, a more precise prognosis may be possible with the use of a score based on gene expression. Moreover, the incorporation of immune-based scores, such as tumour infiltrating lymphocytes, neutrophils, or macrophages, may yield composite scores that can operate as prognostic and predictive biomarkers, opening the door to the development of innovative, individualised immunotherapeutic therapies. Better clinical treatment of colorectal cancer might be made possible by ongoing efforts to develop and confirm multi-gene patterns as predictive biomarkers.

LIMITATION

Mathematical models, equations, and other mathematical expressions provide the backbone of quantitative approaches, which in turn rest on a number of presumptions. The following assumptions are not guaranteed to be true in all situations. Neglecting this caveat might have disastrous results. As quantitative procedures sometimes need the participation of professionals, they may come at a high cost. Even the largest corporations make little use of quantitative approaches since so many applications are not profitable enough to warrant the investment. Managers often rely on gut feelings and precedents rather than cold, hard data when making choices.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY2481

Some errors in quantitative research include insufficient data, conflicting definitions, a poor selection of samples, an improper research design, unsuitable comparisons, and erroneous presentation. Quantitative methods cannot be utilised to analyse qualitative phenomena since they do not take into consideration intangible and non-measurable human traits. Methodologies do not account for intangibles such as a manager's ability, attitude, or enthusiasm. Nevertheless, the procedures might be implemented indirectly by first quantifying previously abstract statements. To determine, for example, a manager's Level of intelligence, it could be necessary to give a certain amount of weight to several qualities.

REFERENCES

[1]. Brenner, S. (2017). THEORIES OF GENE REGULATION. *Laboratory of Molecular Biology*, 1-5.
[2]. Esteban, R. &. (2017). Differential gene expression in disease: a comparison between high-throughput studies and the literature. *BMC Medical Genomics*, 1-10.
[3]. Izum, K. (2016). Disorders of Transcriptional Regulation: An Emerging Category of Multiple Malformation Syndromes . *Molecular Syndromology*, 1-12.
[4]. Jackson, M. M. (2018). The genetic basis of disease. *Essays in Biochemistry*, 1-81.
[5]. Kleinjan, D. &. (2016). Long-Range Control of Gene Expression: Emerging Mechanisms and Disruption in Disease. *Am. J. Hum. Genet.*, 1-25.
[6]. Lee, T. &. (2013). Transcriptional Regulation and Its Misregulation in Disease. *Cell*, 1-15.
[7]. Zhao, J. Y. (2015). Ranking Candidate Disease Genes from Gene Expression and Protein Interaction: A Katz-Centrality Based Approach. *PLoS ONE*, 1-9.
[8]. Abraham, K. (2016). Disorders of Transcriptional Regulation: An Emerging Category of Multiple Malformation Syndromes . *Molecular Syndromology*, 1-12.