Integrative Transcriptome Data Analysis Reveals Psoriasis Signature Genes and Its Potential Role in Drug Repurposing

A DISSERTATION

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
BIOINFORMATICS

Submitted by:

Pawan Singh Gangwar

2K18/BIO/08

Under the Supervision of

Dr. YASHA HASIJA

DEPARTMENT OF BIOTECHNOLOGY

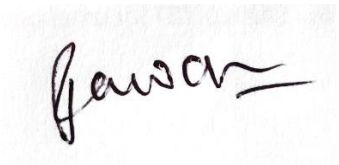
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

JUNE, 2020

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Pawan Singh Gangwar, 2K18/BIO/08 of M.Tech (Bioinformatics), hereby declare that the Project Dissertation titled "Integrative Transcriptome Data Analysis Reveals Psoriasis Signature Genes and Its Potential Role in Drug Repurposing" which is submitted by me to the Department of Biotechnology, Delhi Technological University, Delhi in partial fulfilment of the requirements for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of the Degree, Diploma Associateship, Fellowship or other similar title or recognition.



Place: Delhi **PAWAN SINGH GANGWAR**

Date: 30/06/2020

DEPARTMENT OF BIOTECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "Integrative Transcriptome Data Analysis Reveals Psoriasis Signature Genes and Its Potential Role in Drug Repurposing" which is submitted by Pawan Singh Gangwar, Roll No. 2K18/BIO/08, Department of Biotechnology, Delhi Technological University, Delhi in partial fulfilment of the requirements for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi

Date:

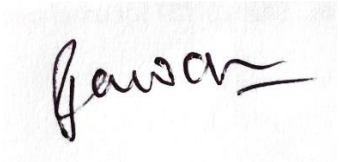
Dr. YASHA HASIJA
SUPERVISOR
Associate Professor
Department of Biotechnology
Delhi Technological University

ABSTRACT

Expression profiling of gene transcripts had been applied in biomedical researches successfully for over a decade. Psoriasis being a chronic inflammatory skin disorder have complex pathological features and unmet pharmacological demands. Therefore, a number of research studies have identified the genes which are differentially expressed in psoriasis skin as compared to the control or normal skin. Although, there is a considerable variance in the differentially expressed gene (DEG) list as reported by several research groups and the precise cause of psoriasis occurrence is still not understood entirely. In this study, the gene expression data from three different microarray studies, consisting of 117 samples in total and more than 1,35,000 transcripts, was analysed with the help of a ranking based approach. Subsequently, the 66 psoriatic gene expression signatures identified in total, were consistently showing dysregulation across the three studies. Furthermore, functional annotation of the identified genes implicated their role in skin development, epidermal development, keratinocyte differentiation, inflammation, immune response, and antimicrobial humoral response. By using a bioinformatics approach, skin development and keratinocyte differentiation pathway were identified as most over represented pathways among 66 signature genes. The main role of keratinocytes is in barrier function due to their presence in the epidermis. An enhanced understanding of keratinocytes function in psoriasis will prove to be important in developing novel barrier therapies for the disease. Finally, a framework is presented which identifies potential drug repositioning opportunities utilising a systemic data-driven approach which helps to discover association amongst genes, diseases and drugs. Such mechanisms facilitated potential drug repurposing for psoriasis by identifying and suggesting new indications of existing drugs.

ACKNOWLEDGEMENT

I wish to express my deep sense of gratitude and indebtedness to Dr. Yasha Hasija, Associate Professor, Department of Biotechnology, Delhi Technological University, Delhi for introducing the present topic and for her inspiring guidance, constructive and valuable suggestion throughout the project work. I am heartily thankful for her guidance and support during the project. Her able knowledge and expert supervision with unswerving patience fathered my work at every stage, for without her warm affection and encouragement, the fulfilment of the task would have been very difficult. I would also like to extend my sincere gratitude towards all the members of the Department of Biotechnology for gratuitously helping me in the successful completion of the project. I am greatly thankful to my family for constantly encouraging me and giving me time and unconditional support while pursuing this work. I am genuinely appreciative of all my friends for their suggestions and moral support during my work. I also convey my heartfelt gratitude to all the research scholars of the Genome Informatics Lab for their valuable suggestions and helpful discussions throughout the course of this research work.



PAWAN SINGH GANGWAR

(2K18/BIO/08)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DEGs	Differentially Expressed Genes
CDR	Common Dataset Ratio
CMap	Connectivity Map
DNA	Deoxyribonucleic acid
EST	Expressed Sequence Tags
FC	Fold Change
FDA	Food and Drug Administration
GAGE	Generally Applicable Geneset Enrichment
GEO	Gene Expression Omnibus
GSEA	Gene Sets Enrichment Analysis
GWAS	Genome Wide Association Studies
HCA	Hierarchical Cluster Analysis
HuGENet	Human Genome Epidemiology Networks
iDEP	integrated Differential Expression and Pathway analysis
KEGG	Kyoto Encyclopedia of Genes and Genomes
NCBI	National Centre for Biotechnology Information
PAGE	Parametric Analyses of Geneset Enrichment
PCA	Principal Component Analysis
RNA	Ribonucleic acid
RNA-Seq	RNA Sequencing

CHAPTER 1 INTRODUCTION

Psoriasis is a chronic inflammatory disease which affects 1 to 2% of individuals, with significant variations in several parts of the world, estimated as eminent as 8% in some geographical areas [1]. The common age group in which psoriasis occurs is 50-69, although it may affect at any age [2]. The average onset age of psoriasis as indicated by some studies was 33 years, while 3/4th instances happened before the age of 46 years [3]. From a long time many studies have focussed on psoriasis, still the etiology of this complex skin disease is not clear, while there has been genetic predisposition evidence [4]. In genetically predisposed individuals, the activation of T-cell occurs in the dermis and epidermis, in response to specific antigens. This leads to a cytokine cascade, which causes keratinocyte proliferation and plaque formation that is characteristic of disease. As presently understood, it is suggested that psoriasis could be a T-cell mediated skin-specific autoimmune disorder, although no responsible self-antigen had been yet discovered [5]. In genetically susceptible individuals, psoriasis could even be stimulated by the internal and external triggers, like infections, stress, sunburn, systemic drugs, mild trauma, or physical injury [6]. These factors results in T-cell activation and cytokine generation, which is followed by neutrophils influx and inflammatory mediator release, leading to cutaneous lesions development.

The skin and nails are involved in psoriasis, and there are number of comorbidities associated with it. Skin lesions developed are generalized or localized, demarcated sharply, consisting of red plaques & papules, and are normally covered with silver/white scales which causes itchiness, pain and stinginess. As approximated, 66% of psoriasis patients develop chronic, psoriasis arthritis (inflammatory arthritis) which causes deformation of joints and disability eventually. And about two-third of all psoriasis suffering patients develop changes in their nails. Psoriasis affected individuals usually report of being at an elevated risk for developing several critical clinical condition like cardio vascular disorders and some other NCDs (non-communicable diseases).

Psoriasis treatment is still depends on curbing the symptoms. Available therapy involves topical, systemic therapies and also phototherapy. These therapies are often used in combination in general practice. Usually aiming at remission, there is a lifelong need for treatment. There is not any therapy so far which can give the hope for an absolute psoriasis

cure. In addition, psoriasis patient care not only requires treatment of skin lesions, joint involvement, but also, it is crucial in identifying and managing common comorbidities which exists already or might develop, which includes metabolic, cardiovascular disorders and also psychological conditions.

1.1 Skin and nails

Psoriasis involves skin, nails, and is also linked with several comorbidities. In addition, psoriatic arthritis might develop in some patients. It is much commonly recognised papulo-squamous disease. Psoriasis clinical features are generalised or localised skin lesions characterised by sharply demarcated, mostly symmetrical, round/oval, red plaque and papules commonly covered with sliver or whitish scales causing itchiness & pain. The symptoms related to psoriasis often include scale formation (92%), itchiness (72%), erythema (69%), fatigue (27%), swelling (23%), burn sensation (20%) and bleeding (20%) [7]. Psoriasis manifests several different forms (Fig. 1). Depending on types, location of skin lesion, onset age of disease and its course, various psoriasis clinical classifications include plaque, intertriginous, guttate, pustular and erythrodermic (Table 1). Nail psoriasis might occur in addition to skin involvement or can occur alone, as the only psoriasis symptom. It is prevalent in 4.2-69% of all the cases of psoriasis [8, 9]. Nail psoriasis not only affects aesthetically, but also restricts manual dexterity. The severity might be varied, acute or chronic, involving intense destruction of a single nail or all nails. These patients suffered from psoriatic arthritis more often. Also, nail psoriasis makes the patients susceptible to bacterial or fungal infections, as reported in 4.6-30% cases [10].

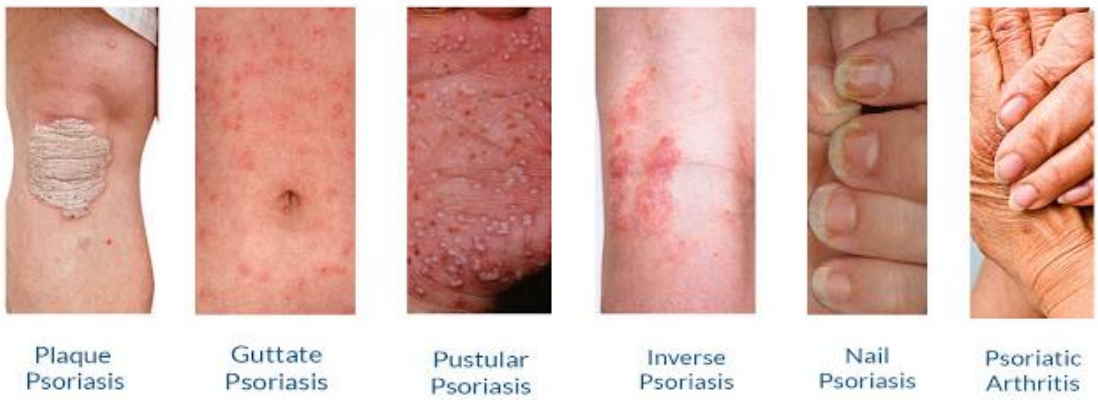


Figure 1. Types of Psoriasis

1.2 Psoriatic Arthritis

Apart from skin, psoriasis could be associated to psoriatic arthritis or inflammatory arthritis involving spine joints and other body joints. Although, it happens without the presence of specific blood antibodies (seronegative spondyloarthropathy and rheumatoid factors). It affects 1.3-34.7% of psoriasis patients [11, 12] with no sex based data. The symptoms of psoriatic arthritis vary, however, dactylitis (swelling in fingers/toes), arthritis in fingers, peripheral arthritis, enthesitis (swelling of tendon-bone conjunction sites) and spondylitis being the most common. It changes physical appearance leading to decreased fitness and causes severe pain. Occurrence of psoriatic arthritis alone is rare i.e. in absence of psoriasis, rather it typically occurs with long standing skin lesions.

Table 1. Psoriasis types & their manifestations

Type	Manifestation	Location
Psoriasis vulgaris (plaque psoriasis)	• Most common • Affects 58-97% of all cases [13, 14] • Plaques are raised, sharply demarcated, red, dry usually covered by white/silvery scale	• Scalp, area behind ears, extensor surface of knees and elbows, sole, palm, nails and face
Inverse psoriasis (in folds and genital region – intertriginous psoriasis)	• Affects 12-26% of all cases [15, 16] • Plaques are flat, sharply demarcated, deep red, wet with scales absent	• Flexural body areas like axillae, inframammary crease, groin, genital area, antecubital and popliteal fossae, gluteal cleft & other body folds
Guttate psoriasis (droplet psoriasis)	• Affect 0.6-20% of all cases [14, 8] • Papules and plaques are reddish, drop-like	• Areas involve arms, legs and trunk. Onset is linked with prior skin symptoms and streptococcal infection.
Pustular psoriasis	• Affects 1.1-12% of all cases [14, 15] • Pustules are coalescent and filled with non-infectious pus	• On entire body after a trigger as a single episode, or on small areas like palm, sole, nails and fingertips
Erythrodermic psoriasis	• Most serious • Affects 0.4-7% of all cases [14, 15] • Fiery redness and leads to hypothermia, cardiac failure and hypoalbuminemia	• Causes exfoliation of most of the body surfaces

1.3 Associated Diseases and Comorbidities

Existence of psoriasis with some chronic systemic disorders is reported by several researchers, of which frequently referred are cardiovascular disorders, metabolic syndrome, diabetes mellitus & dyslipidaemia, hypertension and Chron's disease. There is elevated rate of comorbidities even in children when compared to normal ones [17]. Most of the studies have been discussing the relationship between psoriasis & cardiovascular disorders. There is an increased load of vascular inflammation & subclinical atherosclerosis in psoriasis patients [18]. Also, the serum lipid level (cholesterol and triglycerides) is significantly higher as compared to normal individuals [19]. Stroke and atrial fibrillation is linked with psoriasis, which often aggravates in adolescent patients [20]. Although, it is not sure of psoriasis developing cardiovascular disorder as being an independent risk-factor. However, gain of weight or obesity, smoking tobacco exhibits to be an independent psoriasis risk factor [21]. The measure of the metabolic syndrome, erectile dysfunction and depression is also higher in psoriasis patients [22]. Despite many publications linking comorbidity to psoriasis, the independence and casualty on several diseases remain unclear requiring further studies.

1.4 Purpose of Study

The annual cost for treating psoriasis had been estimated at $11.5 billion in US [23]. However, currently there is no specific/targeted therapy for psoriasis clinically. Present major therapeutic strategies depend on severity of psoriasis at time of presentation and are anti-inflammatory, subsequently leading to reduced turnover of epidermal keratinocyte and flattening of plaque. The 3 major forms of therapy used are topical, phototherapy & systemic [24]. Topical therapy includes creams, ointments, lotion, foams or gels applied to the skin and corticosteroids like hydrocortisone or betamethasone. It usually treats mild psoriasis, and progresses to phototherapy in low response case. Phototherapy uses ultraviolet (UV) light therapy. Systemic therapies are required in moderate-severe psoriasis which includes common first-line drugs like cyclosporine and methotrexate. Various microarray-based experiments are regularly used to investigate psoriasis pathogenesis, which includes DEGs analysis (differentially expressed gene) between psoriasis patients & normal controls. Although, there has been significant variation in DEGs lists reported by several groups, which may have resulted from analytical bias. The approach used here would help to answer the frequency and enrichment of same sets of relevant psoriasis genes/pathways occurring across different datasets.

1.5 Organisation of Dissertation

1.5.1 Literature Review

In this section, a comprehensive summary of the research work done in the past years on transcriptome analysis and differentially expressed genes (DEGs) in relation to dermatological disorders, especially psoriasis, is described. The previous knowledge associated with analysing DEGs and methods of performing pathway and enrichment analysis to establish network analysis assisted in finding the present results.

1.5.2 Methodology

This section presents an analytical approach which depends on consistency, reproducibility, statistical significance and which is biologically relevant, consisting of following steps: (1) DEGs identification based on biological relevance as fold-change (FC), from individual studies [25]; (2) overlap finding between lists of ranked genes across the data sets utilising common data set ratio (CDR); (3) overlapping DEGs ranking based on p-value; (4) discriminating psoriasis from control individual samples using psoriasis signature genes; (5) providing increased understanding of crucial pathways modulated by the therapeutics linked with resolution of disease; and finally (6) describing new potential therapeutic intervention strategy based on drug repositioning approaches. This methodology will allow for better understanding the pathobiology of psoriasis and implementing more precise intervention strategies by identifying potential drug repurposing opportunities for psoriasis. Analysis of multiple publicly accessible transcriptome or gene expression data can be a much effective approach, and even economical for determining disease related genes and pathways [26, 27].

1.5.3 Results, Discussion and Conclusion

This section describes the results obtained from the employment of the mentioned methodology, and interprets and extract the meaning from the results in light of what is already known about the disease research. The conclusion part not only provides the summary of the current work, but also describes the significance of the results, with an open window for the possible future research for the betterment of mankind.

CHAPTER 2 REVIEW OF LITERATURE

The flow of genetic data as stated by central dogma of molecular biology is from genes to proteins or cell functions, which is a 2 step process as follows: 1) DNA (heritable genetic data repository) is transcribed into RNA (a short-lasting data carrier) by the enzyme RNA polymerase; 2) mRNA (messenger RNA), a kind of RNA, is translated into protein. Hence, the complete set of all RNA present in a cell, or population of cells or an organism is known of transcriptome. As all the RNA is not translated into the protein, some of them perform structural function (e.g., rRNA in ribosome assembly); some as transporters (tRNA); some of the others perform regulatory function (e.g., long non-coding RNAs [lncRNAs], short interfering RNA [siRNAs]) [28]. Nonetheless, the non-coding RNAs frequently play role in human disorders like cardiovascular, cancer and neurological. Transcriptomics although, most commonly applies to the coding transcripts i.e. mRNAs, it also provides valuable information regarding the noncoding RNAs of the cell.

2.1 Transcriptome Analysis

Technologies used in analysis of transcriptome, states the methods or techniques which are used for studying transcriptome i.e. the sum of all RNA transcripts of an organism. Transcription express the data which is stored in the DNA of an organism's genome. In transcription, the coding RNA (mRNA) acts as a transient intermediate molecule, while the non coding RNAs serves several other broad functions. As summed up, a transcriptome clicks a snapshot of the overall transcripts existing in the cell in time.

Measuring expression of genes of an organism in multiple tissues, states, or points of time, gives data on how the genes are regulated & also uncovers the detail of an individual's biology. This also helps in inferring already unannotated gene functions. Transcriptomic analysis has enabled examining changes of gene expression in various life forms and had been instrumental in the comprehension of human disorders. A whole analyses of gene expression permits detecting expansive coordinated patterns that could not be recognised by many targeted assays.

2.2 Uses of Transcriptome Analysis

The analysis of transcriptome is usually done to compare specific sets of sample or tests, much common are healthy/control & disease/test conditions. The differences might be because of various external environmental conditions like hormonal impacts or toxins. Tissues at different developmental steps could be characterised molecularly. Like the stem cell transcriptome helps to understand the procedures of development of an embryo or cellular differentiation. In cancer, transcriptome analysis states classification, process of pathogenesis & also outcome prediction. These researches could group cancers past anatomy area & histopathology. Gene-based benchmarks could be established by outcome predictions to foresee tumour prognosis and treatment response. Such methodologies are as of now being used for precision medicine, cancer personalised therapies. Due to its much wide approaches, analysis of transcriptome is an extraordinary source for identification of therapy targets.

2.3 Methods of Transcriptome Analysis

The study of whole transcriptome started in mid 1990s, and technical innovations since the late 1990 had made transcriptomic a comprehensive discipline. The beginning attempt to capture a fractional human transcriptome was published in 1991, reporting 609 human brain mRNA sequences [29]. Two human transcriptomes were published in 2008, which consisted of millions of sequences derived from transcripts which covered 16,000 genes [30][31], and till 2015, transcriptomes for 100s of individuals had been published [32][33]. Nowadays, transcriptome are routinely generated for distinct disorder forms, tissue, and also individual cells [33][34].

Transcriptome study had been characterised by repetitive technical advances which transforms the discipline. Producing RNA transcript data could be accomplished by two key technologies in the field: microarrays (transcript hybridization to an arranged array of nucleotide probes), which quantifies a set of predetermined sequences; & RNA-Seq (RNA sequencing), which utilises high-throughput sequencing to catch all sequences.

2.3.1 Microarrays

Microarray comprises of short nucleotides oligomer i.e. probes that are arrayed on a hard glass like base or substrate [35]. Hybridization of fluorescent tagged transcripts to these probes

determines abundance of transcripts. The fluorescent intensity at every probe's area on array shows the transcripts ampleness for that specific probe's sequence [36]. The prerequisite for microarrays is that it requires several prior understanding of the organism of interest, such as, in the form of ESTs (expressed sequence tags) or an annotated genome sequence library which could be utilised to achieve the array probes. Fig. 2 provides a summary of DNA microarrays [37].

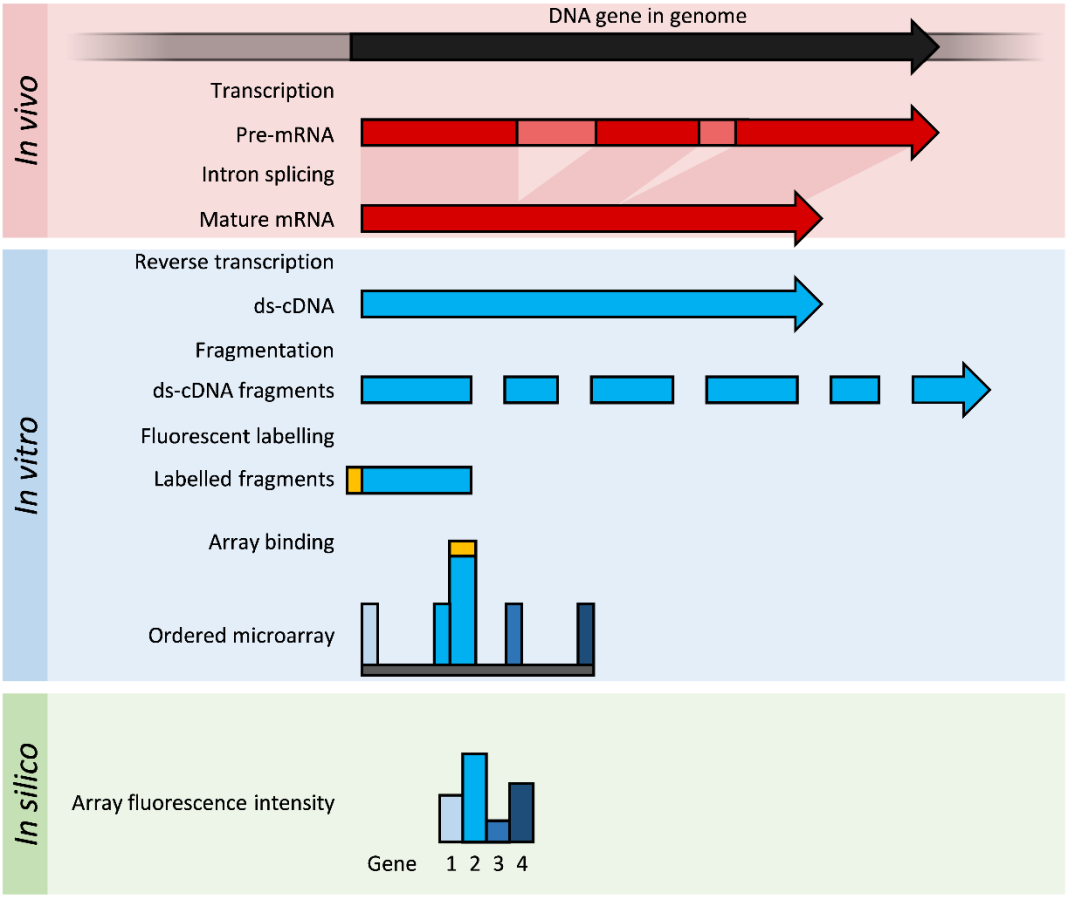


Figure 2. Microarray

2.3.2 RNA-Seq

RNA-Seq is amalgamation of high-throughput sequencing methods with computational techniques for capturing & quantifying transcripts obtained in RNA extract [38]. The nucleotide sequences obtained are usually of 100bp length, still could range from 30bp to above 10,000bp which depends on used sequencing technique. RNA-Seq gives the advantage of sampling transcriptomes deeply by several shorter fragments from a transcriptome to permit computation reconstruction of the primary RNA transcript by arranging reads to a reference genome or de novo (to each other) [39]. The characteristic RNA-Seq magnitude of 5 orders is

a main leverage over microarray transcriptome. Additionally, the amounts of input RNA are very low for RNA sequencing (nanograms) when correlated to microarrays (micrograms), allowing detailed examinations of cell structure, brought downwards to the level of individual cell when joined with linear amplified cDNA [40]. Hypothetically, there is not any above quantification limit in RNA-sequencing, & the background signal is also much lower for 100bp read in the non-repeatitive regions [41]. Fig. 3 provides a summary of RNA sequencing [37].

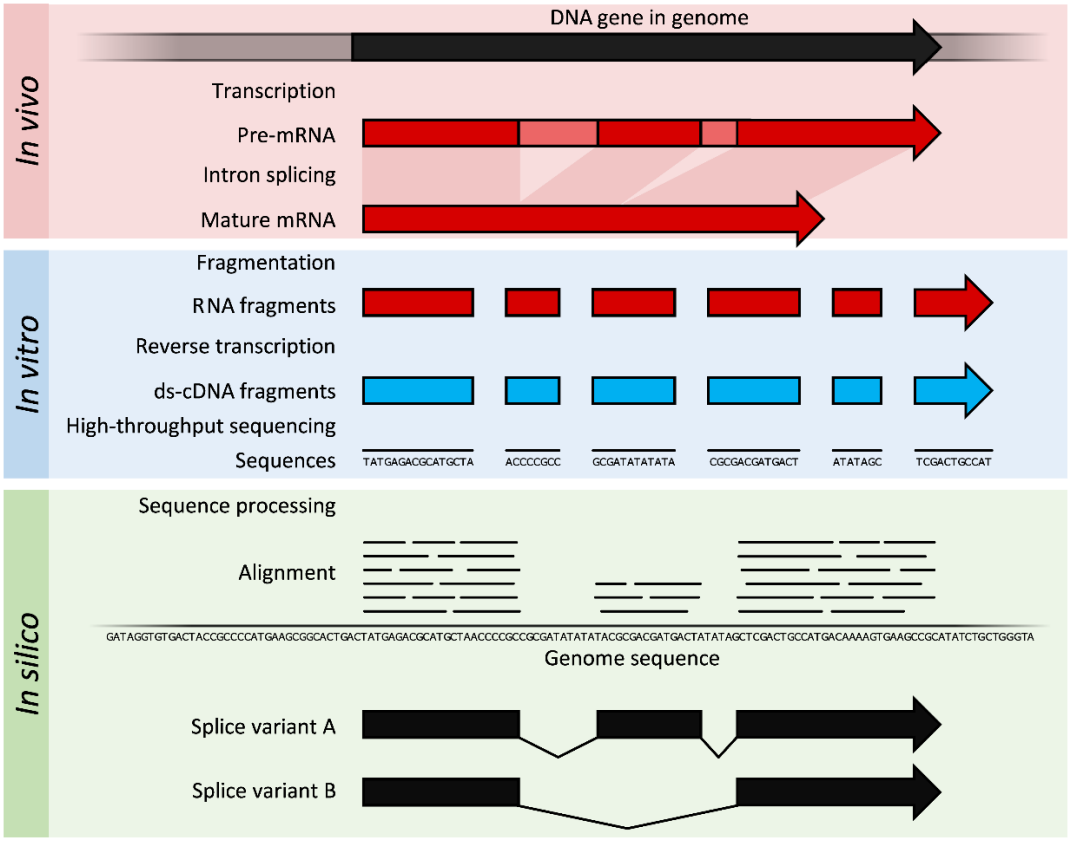


Figure 3. RNA Sequencing

2.4 Data Analysis

Improved sequencing methods demanded improved analysis of data to manage the expanded volume of information created by every transcriptomic experiment. Subsequently, the results are stored in transcriptomic databases, serving as necessary tools for transcriptomic analyses. For instance, GEO (Gene Expression Omnibus, NCBI) stores millions of experiments of transcription profiling. The possible utilizations of such data is beyond the original objectives of an analysis. Usual output includes transcript level quantitative tables. This needs definite experimental algorithm, frequently distinct to the utilised strategy. Also, software packages

connect information from divergent procedures to recognize group of identical expressed genes, or functionally significant differentially expressed metabolic and regulative pathways.

2.5 Presentation (Differential Expression)

The output of transcriptome analysis frequently appear graphically in form of heat maps, a colour-code method which represent distinct expression levels of given genes in distinct samples (Fig. 4) [37]. These presentation often show sample clustering, which help in identifying same gene expression samples. Every column stores the value of differential gene expression for only one sample. The relative gene expression is shown with help of colors; red (highly expressed), white (median expressed) and blue (low expressed). The samples could be of distinct individuals, tissue, environment or even health condition.

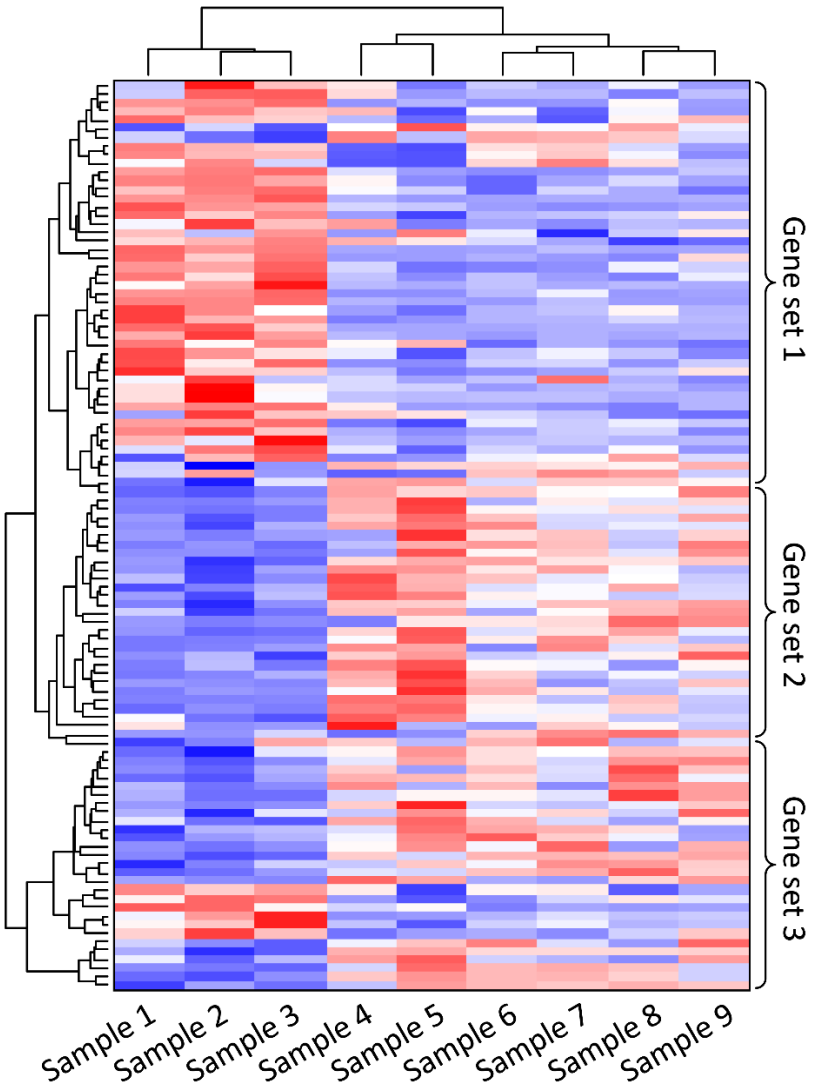


Figure 4. Heatmap (Gene co-expression across different samples)

2.6 Transcriptome Analysis in Human Skin

Skin, being accessible conveniently, was amongst the early target analysed utilising 'omics' and very soon the methods were embraced by dermatology [42]. A classical case of coordinate transcription regulation was seen in the cultured fibroblast after stimulation by serum [43]. Addition of serum not just causes fast resumption of cellular cycle, yet also a characteristic wound curing response, a physiologic act of fibroblasts in wound healing [44]. Transcription response of epidermis keratinocyte to ultraviolet light, infection, vitamin, hormone, inflammatory and immune-modulating cytokine, allergens & toxins had been characterised, and also the change linked with epidermis differentiation [45] [46].

The gene signature which characterise different cell type in human skin, were utilized to characterize 20 particular expression signature, which includes those for keratinocyte, adipocytes, endothelium, melanocytes, immune cell, hairs follicle, sweat, sebaceous & apocrine gland. It catered a platform called as SkinSig, which was then utilised to analyse 18 distinct skin states, giving in contextual depiction of, e.g., immune cell influx in differentiation or inflammation variations in cornified diseases [47]. There are several tools for transcriptome analysis e.g., AltAnalyze, iDEP, etc. iDEP (integrated Differential Expression & Pathway analysis) an easy to use, online interactive tool used for exploratory analysis of data, differential expression, and pathway analysis.

Swindell et al [48] analysed existed sets of data for evaluating genome-wide expression in lesion from163 individuals affected with psoriasis with the aim for identification of systems which drive differential expression and characterising heterogeneity amongst lesion in such vast sample. They classified 1233 up-regulated DEGs & 977 down-regulated DEG. Up-regulated DEGs were associated to 56% keratinocyte activities, 14% lesion infiltration by Tcells and 11% macrophage. In contrast, down-regulated DEGs were attributed/linked with 63% adipose tissues, 14% epidermis and 4% dermis. Epidermis and keratinocyte DEGs showed enrichment of genes induced by IL-1, IL-17A & IL-20 cytokine, and even showed disproportionate association with binding sites of AP-1. Amongst all the patients, an increased inflammation signature was exhibited by 50%, with heightened gene expression shown by T-cell, monocyte & dendritic cells. IFN-γ-strong signature was displayed by 66% of patients, with heightened gene expression induced by IFN-γ, additional to various other cytokine (such as TNF, IL-1 & IL-17A). These differences/patterns seen in gene expression could be utilised as biomarkers for differentiating between etanercept responders & non-responders.

2.7 Drug Discovery & Connectivity Map

The CMap (Connectivity Map) uses the transcriptomes & uses gene expression profiles as a typical communication to link biology-chemistry & clinical condition for deciphering disorder-gene-drug associations paying little heed to the microarray platforms utilised [49]. The methodology starts with a phenotype of intrigue, for example, a biological state or disorder to infer a priori stated gene expression signatures, which is, differential expressed gene set which are unique and represent to the phenotype addressed. A signature gene set consists of both significant up-regulated & down-regulated gene & are crowded by far less than overall no. of transcripts, usually in amount of little dozens & hundred. Gene signature set given could after that utilised for query & make comparison against the vast referenced catalog of gene expressions profile resulted after the drug or other perturbation cell line treatment.

The public funded CMap references catalog earlier consisted profile of 164 small molecules and afterwards increased to 1309 F.D.A. approved drugs. Such smaller molecule are tested in 5 human cell line, producing above 7000s expression profile in database of CMap. This even serves an interactive internet site & an on-line tool for conducting CMap query againsts the chemical references catalogue. With the help of an easy but productive pattern match algorithm which depends on rank-ordered non-parametric Kolmogorov–Smirnov statistic [50], the similarity metrics between a test gene signature & all references datasets gets transformed to a connectivity score which ranges from -1 to +1 and reflects nearness/association between the expression profiles. A +ve score denote similarity degree (positively correlated) & a -ve score denote an inverse similarity (negatively correlated) between a signature query & a reference profile obtained from an individual chemical perturbations, hence implying the exposure to a specific chemical could exasperate (potential inducers) or reverse (potential therapeutics) the expression patterns of phenotypes of interests.

Xiaoyan et al [51] presented a model to develop new therapeutical psoriasis intervention strategy by using publicly accessible clinical transcriptome datasets. Thus, exploring the described molecular models for psoriasis, effect of consequent perturbation of such models by drug molecules & an integration analyses, they proposed a disease signature of psoriasis, identified potential drug repositioning opportunity & presented new methods for selection of targets. They anticipated that the methodologies outlined or alike approach would further support biomarkers discoveries and new drug development for psoriasis.

CHAPTER 3 METHODOLOGY

3.1 Strategy for Data Analysis

For exploring gene and related pathway associated to psoriasis, gene expression analyses was performed utilising psoriasis data sets obtained from three independent studies [52] [53] [54]. Firstly, each dataset originated from the public available database was analysed for determining DEGs (differentially expressed genes) having fold change FC ≥1.5 [55]. Then, the DEGs which were present consistently in datasets individually were found and ordered based on p-value [55]. Then the obtained DEGs were grouped based on their function for identifying the most affected psoriasis related pathway. Fig. 5 summarises the data collection and analysis steps.

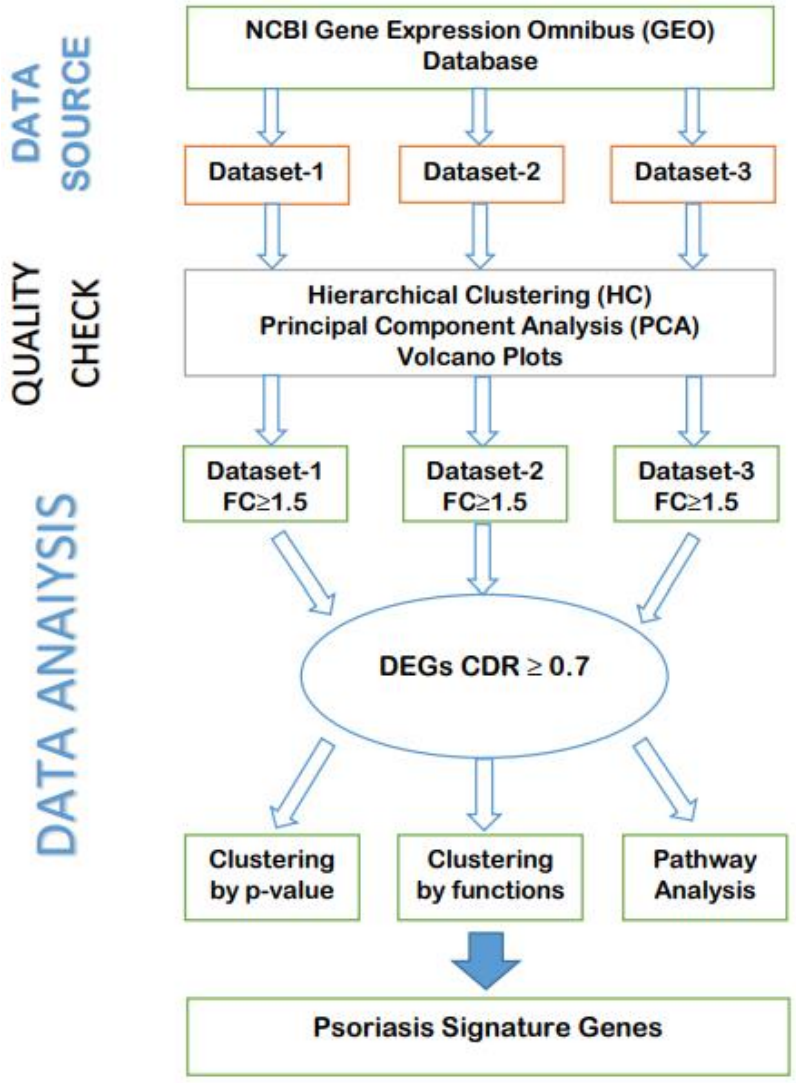


Figure 5. Steps of Transcriptome Analysis

3.2 Identification of Psoriasis Gene Expression Datasets

Gene Expression Omnibus (GEO, NCBI) database is utilised to identify psoriasis microarray datasets [56, 57]. Relevant literature published between 2015 and 2019, analysed for DEG in psoriasis vs normal subject was selected. The following inclusion criteria were followed for selecting eligible datasets: a) the data set essentially compares psoriasis patient with healthy (non-psoriatic) control, & b) the data set should be generated from similar type of tissue like skin. From every study, the data which follows were extracted: (1) GEO accession numbers, (2) type of sample, (3) platform, (4) no. of psoriasis & non-psoriatic individual, and (5) gene expression value; as shown in Table 2.

Table 2. Datasets used for current study

Data	GEO ID	Sample Type	Platform	Sample Size	References
1	GSE78097	Skin	Affymetrix (GPL570)	33	[52]
2	GSE47751	Skin	Affymetrix (GPL570)	34	[53]
3	GSE80047	Skin	Affymetrix (GPL570)	50	[54]

3.3 Data Analysis

Human psoriasis microarrays data sets that passed the inclusive criterias were downloaded from GEO, NCBI database. Three independent genes expression microarrays studies, consisting of an overall 117 sample size were utilized with greater than 135,000 transcripts which represented approx. 5,000 unique gene (depending on Unigene cluster). Tables of data consisting of gene expression value were constructed using GEO2R, having gene row and experiment/sample in column [57]. GEO2R is an interactive online tool which processes tables of data utilising the GEOquery [58], and limma R package of Bio-conductor project [59]. GEOquery R-package was utilised to send GEO data into R data-structure which could be utilised by different R-packages. It manages broad ranges of experiment design and types of data and employs several test correction on p-value in helping correct for the occurrences of false positive. Benjamini & Hochberg FDR (false discovery rate) was used as it is the most usually utilised adjusting methodology for microarrays data & gives a better balance between discoveries of statistical significant gene and false positive [60]. Transcripts occurring in atleast 2 of 3 data sets (CDR $\geq$ 0.7) were found and arranged in accordance with their average FCs. Un-supervised hierarchical clusters analyses (HCA) & principal components analyses (PCA) were done with data received from psoriasis & non-psoriasis group with the help of the web

application iDEP (integrated Differential Expression & Pathway analyses) to detect outliers [61]. Fig. 7 shows FDR adj. p-values vs FC volcano plot of GSE78097 data set for checking the initial quality of data.

iDEP is a user-friendly web-based tool which uses broadly utilised R or Bioconductor package for analysing data of gene expressions. For EDA (exploratory data analysis), it performed hierarchical cluster analysis, *k*-mean cluster analysis, and PCA (principal component analyses). iDEP uses *limma* and *DESeq2* packages for detecting differential expression of genes. For the group of mutual gene expression, iDEP identify enriched gene ontology (GO) and also transcription factors binding motif in promoter sequence. Pathway analyses is performed utilising GAGE (Generally Applicable Geneset Enrichment), PAGE (Parametric Analyses of Geneset Enrichment), GSEA (Gene Sets Enrichment Analysis), or Reactome PA (Reactome Pathway Analyses). Gene expression data is visualised on KEGG pathways diagram using path viewer.

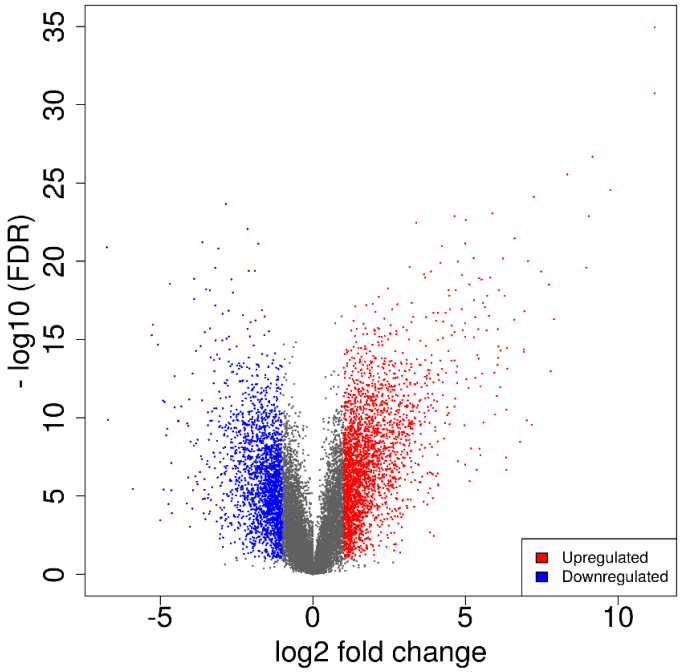


Figure 6. Volcano plot of GSE78097 dataset

3.4 Selection of Signature Genes from DEGs

For subsequent analysis the considered transcript were with fold change FC >1.5 & CDR ≥ 0.7. At this part, in the case where individual DEG was linked with more than one Affy IDs

(affymetrix identity), the highest fold change ID was considered. This yielded 66 genes/transcripts (66 DEGs) which were consistent up-regulated & down-regulated in psoriasis in comparison to control in every dataset. Such gene were utilised for subsequential statistics analyses, functional annotations, pathways, networks and over representation analysis. For checking statistic significance, transcript of every datasets were arranged by p-value (smaller to higher) & considering top 5.0% of gene most relevant in every datasets. For instance, data set GSE6012 has data for a complete 45,118 transcript, so 2256 transcript rank above 5.0% significant levels (45118 x 5/100 = 2256), that includes both up-regulated & down-regulated gene. Utilising the p-values or q-value (adjusted p-value or FDR) doesn't changes the orders of such arrangements appreciably. Table 4 & 5 in Appendix 1 shows down- and up-regulated 66 DEGs.

3.5 Strategy for Drug Repurposing in Psoriasis

The obtained psoriasis signature genes provide a chance for developing integrative analysis for identifying target which may offer probable or potential repositioning opportunity for existing molecule (i.e. drug repurpose) [62]. The outlined methodology includes tractability evaluating by small molecule or biologic like if genes or gene products can be modulated therapeutically with small molecules (druggables) or biologics (biopharmables). GWAS dataset could point loci showing enrichment for genetics validated human target [63] [64]. The genetics associations of such gene with psoriasis was also evaluated utilising HuGENet (Human Genome Epidemiology Networks). It is a loaded knowledge base for disorder genetics association depending on GWAS, primary study and meta analyses evidences [65]. In addition, the pharmaceutical industry pipeline was examined for determining if specific targets have already been explored of either psoriatic or another indication as in Pharmaproject list, which is a leading dataset source of intelligent tracking drug of globally (both approved or in developing phase) in R&D containing updated drug profile in detail, like targets, therapies and development timeline.

Genome-wide expression profiling is often used for studying the effects on cells caused by small molecules, and pathways modulated by them. Understanding consequence of such modulations, pathways, and their interactions is not complete, in context of various cell types of whole organism. CMap which is a novel pathway independent approach ad employs gene expression profile is used in this study for drug repurposing. L1000FWD (Fireworks Display) is used which provides interactive large scale visualization of drug and small molecule induced

transcriptomic signatures. The Clue platform's 'Repurposing' tool is used to find potential drug repositioning opportunities for improving disease treatment.

3.6 Drug Repurposing

Fig. 3 describes analysis pipeline and therapeutic strategy proposed for psoriasis. Out of the 66 psoriasis gene signatures, industry broad pipeline analyses through examining Pharmaproject databases revealed 47 of those gene which had already been considered as target by drug-discovery program. Human genetics association evaluation by psoriasis & the industry broad pipeline of drug discovery program, 12 targets were identified which can be searched for psoriasis utilising molecule which already exists for such target. From 12 targets, 4 are currently being developed for psoriatic indication. Those targets are IL8 (Interleukin 8), CCL2 or MCP-1 (monocyte chemo-attractive protein 1), SOD-2 (Super oxide dismutase 2), and PRKCQ or PKC-θ (Protein Kinase C Theta). Drugs targeted eight remained genes for the indications any different than psoriatic and have reported human genetics association in psoriasis.

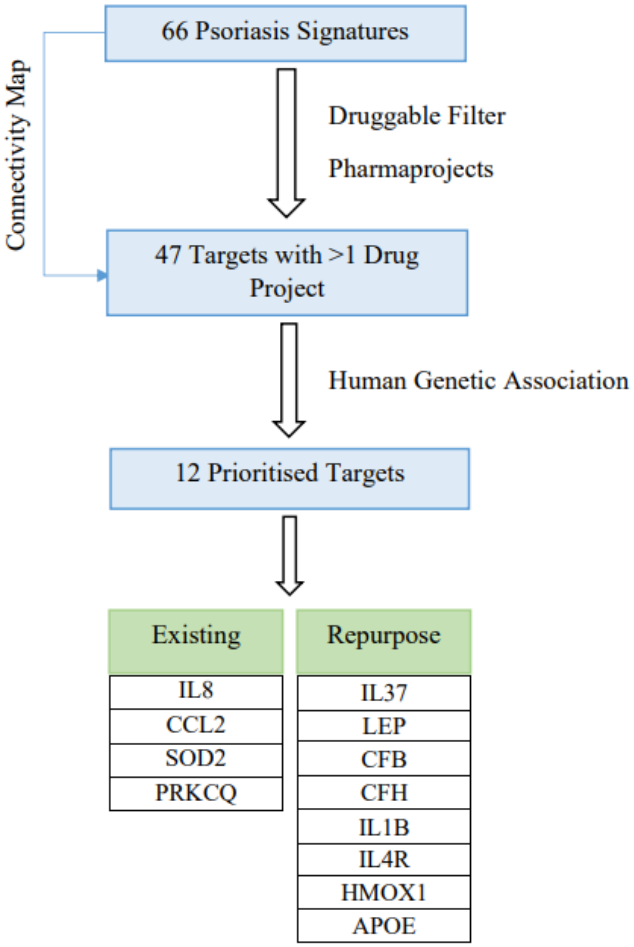


Figure 7. Psoriasis Therapeutic Analysis and Drug Repositioning Strategy

Targeting drugs at such genes product can be assessed for psoriasis correspondingly to the methodology utilized by Sanseau P. et al. [66]. As an illustrating model, IL1B-mRNA & proteins expression level were raised in psoriasis skin lesion [67]. IL-1B comes under potential pro-inflammatory cytokines, & therapy aiming at its modulation can have promising role in autoimmuno inflammatory disorders [68]. It adds in inhibiting insulin dependent keratinocytes differentiation & upgrades proliferated keratinocytes, both of them are psoriasis characteristic [69]. One of the late studies showed that the IL-1B induced gene in cultured keratinocyte were also likewise for showing raised expression in psoriatic lesion. IL-1B level decreases in psoriatic patient's tissue fluid acquired from micro dialysis following clinical therapy [69]. Furthermore, curcumin, a turmeric phyto-chemical, eased imiquimod-induced psoriatic skin like inflammatory skin in a model of mouse, promisingly from inhibiting IL6 & IL1B [70]. Such discoveries provide base for developing or repositioning of existing IL1B modulator that can introduce new psoriasis treatment improvement opportunities. As another model, HMOX1 (haeme oxygenase [decycling] 1), an enzyme that degrades haeme expressed in normal skin keratinocyte & several other tissue, represented a crucial anti-inflammatory and defence mechanisms against the oxidative stress. In transplant injection, drugs targeting HMOX1 are presently being investigated, yet not in psoriasis. HMOX1 shows intense up-regulation in the psoriasis lesion skin [71]. In spite of the fact that the definite mechanism of HMOX1 in psoriasis is still unclear, its contributive function in psoriatic skin anti-oxidant networks is much well present in documents [72], which supports its potentiality as a test-able drug repositioning psoriatic target.

CHAPTER 4 RESULTS, DISCUSSION & CONCUSION

4.1 Results

4.1.1 Differentially Expressed Genes and Signature Genes

The transcripts having FC >1.5 & CDR ≥ 0.7 were conceived from all three GEO datasets. In the cases in which individual DEG were linked with more than one Affy IDs (affymetrix identity), the highest FC ID was considered. This yielded 66 genes or gene expression signatures (GES) which were up-regulated & down-regulated consistently in psoriasis in comparison to control in all data sets. The DEGs which exhibits largest average FC value are: SERPINs (SERPINB4; encodes inhibitory protein for serine proteases enzyme), S100s (S100A7A; encodes S100 Calcium Binding Protein A7A), TCN1 (encodes transcobalamin-1 protein to protect acid-sensitive vitamin B_{12}), BTC (beta-cellulin), c1orf68 (chr 1 orf 68; encode a skin specific protein called LEP-7), CXCL8 (encodes interleukin-8 protein in humans; interleukin 8 is one of the chemokines generated by epithelial cells) & aldo keto reductase (AKR1B10; inflammation related). Innate immune function related DEGs (betadefensin, microsemino protein) & cytokine, chemokine and adhesion responsible molecule (CCL17, CCL18 & CCL22) were seen. Table 3 mentions top consistently up & down regulated genes with respective avg. log FC values.

Table 3. Top up-regulated & down-regulated genes of psoriasis

Top 10 up regulated genes		**Top 10 down regulated genes**	
Gene	**Average FC**	**Gene**	**Average FC**
S100A7A	7.7292619	WIF1	3.967766
SERPINB4	7.2403485	CCL27	3.434093
TCN1	6.0285633	PM20D1	3.161581
S100A12	6.1465832	THRSP	2.398755
SPRR2C	5.6754507	TPPP	2.385762
S100A9	5.6818349	GSTA3	2.355616
GDA	4.6583229	FADS2	2.028950
CXCL8	3.2690672	PIP	1.955895
OASL	4.2746268	KRT79	1.936349
VNN3	3.6742317	APOC1	1.766915

4.1.2 Statistical Analysis of DEGs supports the Gene Signatures

Fig. 8 shows un-supervised HCA (hierarchical cluster analysis) of the datasets. It depends on the mean for total psoriasis individual and control individual was utilised for obtaining similarity amongst individuals in accordance with the correlation measure across all expression value of 66 genes data sets. Dissimilarity is represented by branch height. It should be noted that psoriasis sample differs in length of branch from control. In HCA, Euclidean distances and complete linkage was utilised for obtaining similarity/dissimilarity amongst set of individuals according to the gene expression values of all the 66 DEGs. The passed sample were grouped into two clusters as expected: psoriasis patient vs control individual.

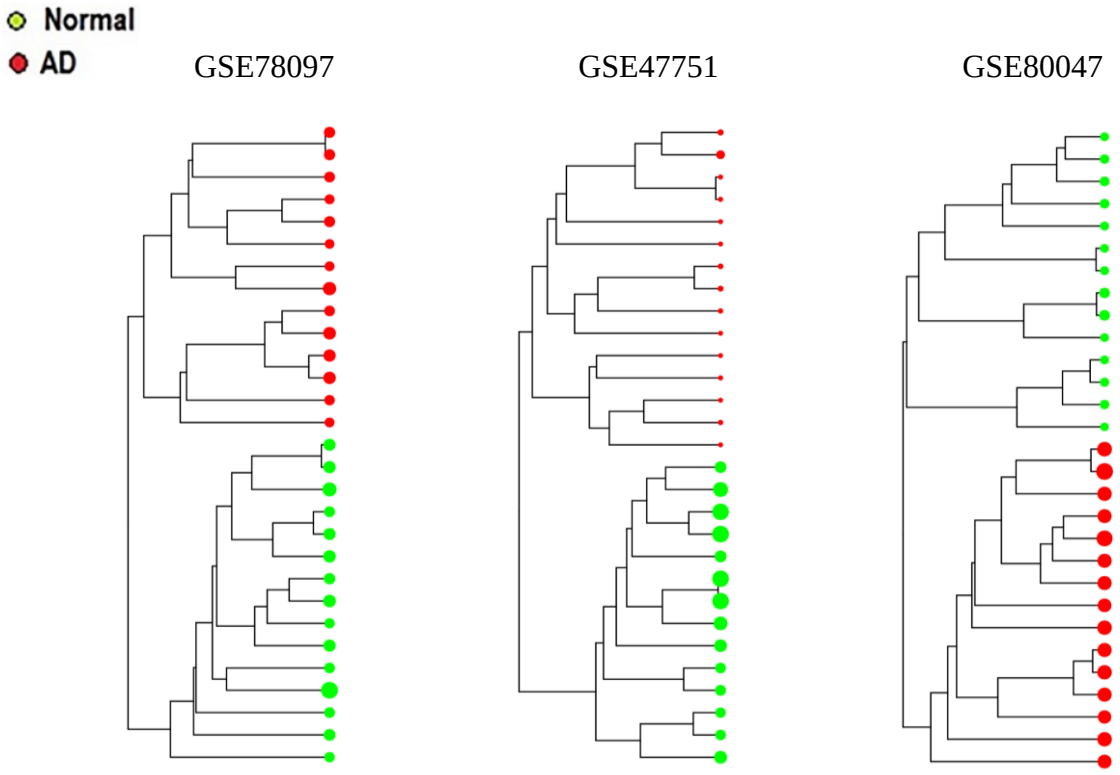


Figure 8. Hierarchical clustering of all three datasets

4.1.3 Functional Annotation and Enrichment Analysis

Functional annotation clustered 66 DEGs into 4 major functional categories as follows:

Cluster A: Genes related to skin development & barrier functions. The peak barrier functions or structures related gene identified: KRT16 (cyto-keratin), COL6A6 (collagen), LCE2B (late cornified envelope), FLG (filaggrin), SCEL (sciellin) and AQP9 (aqua-porin).

COL6A6 and KRT16. These were upregulated in psoriasis skin whilst the rest were downregulated. Moreover, a several genes of such group were significantly linked to the skin development and keratinocytes differentiation pathways.

Cluster B: Genes related to inflammation. The differential regulated cytokines linked gene were IL1F7 (also known as IL37 (Interleukin 37), is an anti inflammatory cytokine and belongs to interleukin-1 family), IL27RA (Interleukin 27 receptor, alpha), chemokine CCL17, CCL18 & CCL22, growth factor EREG (encodes epiregulin protein), SELE (selectin E). IL1F7 expression and growth factors were down-regulated in all of the datasets. IL27 receptors were upregulated in psoriatic datasets.

Cluster C: Genes related to anti-microbial response. The differentially regulated genes were calcium binding protein encoding S100 family gene S100A9 (also called as MRP-14 (migratory inhibition factor related protein 14) or calgranulin B), S100A12 (also known as calgranulin C), and LCN2 (Lipocalin 2). Down regulated gene linked to innate immunity in psoriatic data sets was seen, which were LTF (lactotransferrin), MSMB (micro-seminoprotein ß), & SCGB-2A1 (secreto-globin 2A1). DEFB4 (defensin, ß 4) which encodes ß-defencin 2 (BD2) also called as skin anti-microbial peptide 1 (SAP-1) was upregulated in psoriasis patients.

Cluster D: Genes related to protease and its inhibitors. Differential expressed, top proteases and proteases inhibitor gene recognised were: Serpin family (SERPINB-3, SERPINB-4 & SERPINB-7), PI (proteases inhibitor epidermal), CORIN (membrane bounded serine peptidases), KLK5 (Kallikriens; serine proteases like trypsin), ASPRV1 (aspartic peptidase), TMPRSS4 (transmembrane serine protease) and CTSL2 (Cathepsin). In these, epidermal protease was up regulated in psoriasis and proteases inhibitor was down regulated.

4.1.4 Pathway Analysis

Analysis of 66 DEGs with the iDEP (integrated Differential Expression & Pathway Analysis) revealed that the skin development and keratinocytes differentiation pathway were the most enriched pathway significantly, with FDR (adj. p-value) of 1.14E-13 and 1.69E-12 respectively. The gene linked with the skin development pathway (14 of 66 gene signatures) and keratinocyte differentiation pathway (12 from 66 signature genes) are mentioned in pathway enrichment Table 6 given in Appendix 2. A number of such gene are involved in barrier functions [73]. These results intensely indicates that psoriasis is linked with defects in

the keratinocytes differentiation pathway that is consistently down regulated with terminal differentiated protein. The causative genes of the keratinocyte pathway, arranged by the count of the related published studies, are LCE3D, SPRR3, SPRR1A, TGM1, CNFN, WNT5A, SPRR2G, S100A7, DSC2, KRT6B, SERPINB13, and KRT6A.

4.1.5 Results from CMap Analysis

After CMap was introduced in 2006, it had offered an in-silico screening methodology for identifying gene-drug-disease linkages which could be of application to therapeutic researches [49]. CMap analyses was conducted on the connectivity map data base of the Broad Institute, utilising the psoriatic signature genes. All molecules are assigned connectivity score after CMap analysis, scores ranges from +1 to -1, which reflects nearness/connection between the expressions profile. Positive value denote potential disease (phenotype) inducer, and negative value denotes potential therapeutic molecule. And a list of small molecules were identified (with opposite/negative similarity) that can be assessed for potential therapeutics of psoriasis. Table 7 provides list of 10 potential drugs out of top 50 in Appendix 2. Methotrexate, amongst them is known psoriatic therapeutic, which ranks amongst the top 10 molecules with -ve connectivity score to the psoriatic signatures (potentially therapeutic), which provides further aid for the use of the psoriatic disease signatures. Doxycycline inhibits proliferation of Tcell & production of cytokine, chemokine by human peripheral blood mono-nucleated cell. Parthenolide, an active component of feverfew (*Tanacetum parthenium*), is being used as anticancer agent. The proliferation is inhibited by it and it kills several tumour cells chiefly by inducing apoptosis. Tiabendazole (thiabendazole or mintezol) is already being used as antifungal drug, inhibit the helminthic specific fumarate reductases that suppresses microtubules formation leading to defective glucose uptake. Resveratrol, a characteristic plant phenol, additionally showed up as the top hit having -ve connectivities score to the psoriatic signature genes. It had been accounted to bear numerous probable positive effect in the skin, for example, photo protectivity, wounds healing & preventing tumor of skin. With the discoveries such as plant phenolic inhibits neutrophils elastases, that is available in inflammed tissue & psoriasis lesion [74], resveratrol and/or its derivative may deserve advanced examination to investigate its role in psoriasis.

4.2 Discussion

Gene expression signature are mostly employed as biomarkers for classifying patient from healthy control. As an early analyses step, all psoriasis sample were considered, without regarding their clinical subtypes (mild/severe), & then compared to controls. DEGs & functionally linked categories like cell differentiation, lipid metabolism, microbial defense, epithelium development, skin development, epithelial cell differentiation, keratinocyte differentiation, response to external biological stimulation, response to bacteria, antimicrobial humoral responses, leukocyte mediated immunity, cornification, activation of myeloid cell involved in immunological response, response to fungus, & anti-microbial humoral immunological response mediated by anti-microbial peptides were identified using combined samples analysis. The current study also identifies main psoriasis functional cluster, signature genes & enriched pathway relevant to classification of psoriasis & etiology, such as seeping skin barriers and trans(across) epidermal loss of water associated gene were down regulated in psoriasis. Microbe infection, innate immunity gene, barrier dysfunction and inflammation might be caused by dysregulation in proteases and protease-inhibitors homeostasis. The framework for identifying drug repurposing opportunities showed promising insights by using psoriasis gene signatures. The small molecules obtained from CMap may deserve further research for examining their role in psoriasis.

4.3 Conclusion

Using publicly accessible gene expression data of psoriasis, 66 gene signatures were identified and the most enriched pathways in the datasets were found to be the skin development pathway and keratinocyte skin barrier pathway. Performing functional annotation of epidermal structure genes indicated that tight junction barrier and strata corneum of patient skin are impaired in psoriasis, which is also connected with raised inflammation, irritants or allergen exposures, and microbe infection. It may be because of the epithelial genes and their differential expression, which controls barrier integrity, inflammatory response, metabolism of lipids, and innate immunity. Hence, therapeutics aimed at strengthening barrier integrity, instead of just suppressing inflammation may show potential outcomes in psoriasis treatment. Presently, the psoriasis disease is managed by depending, to a great extent, on avoiding allergens, applying moisturisers, cortico-steroids, immuno suppressant, with clinically no targeted therapies in usage. The role of epidermal protein genes linked to psoriasis, which includes S100A7A and keratinocytes differentiation pathway identified here may be greatly useful for designing novel

therapeutic targets such as targets for barrier therapies in psoriasis. The 66 gene signatures identified, can differentiate psoriasis patient from normal controls, and could also function as biomarkers for therapeutical stratification of psoriasis.

The differentially expressed psoriasis gene signatures provided a chance for developing integrated analysis in identifying the targets which might be accounted by novel drug discovery (discovery of targets) programs and could present promising repositioning opportunities for the existing drug molecules (drug repurposing). The genes obtained after applying CMap analysis and human genetic association were targeted by drugs for denotation apart from psoriasis or drug repositioning. Subsequently, the drugs targeting such gene products can be assessed for psoriatic similarity. Repurposing of existing target modulators responsible for inflammation, can exhibit novel developmental chances for psoriasis therapeutics.

APPENDICES

Appendix 1. Supplementary Data of Down-regulated & Up-regulated Genes

Table 4. Down-regulated genes of 66 DEGs with avg. FC and lowest p-value among the three data-sets.

Gene	Avg. FC	Lowest p-value among 3 studies
WIF1	-3.967766	1.36E-10
CCL27	-3.434093	2.35E-13
PM20D1	-3.161581	0.000359
THRSP	-2.398755	1.71E-05
TPPP	-2.385762	2.20E-15
GSTA3	-2.355616	7.13E-13
FADS2	-2.028950	7.06E-06
PIP	-1.955895	0.00116
KRT79	-1.936349	6.55E-05
APOC1	-1.766915	1.47E-05
SCEL	-1.723102	1.11E-08
RORA	-1.643489	4.97E-10

Table 5. Up-regulated genes of 66 DEGs with avg. FC and lowest p-value among the three data-sets.

Gene	Avg. FC	Lowest p-value among 3 studies
S100A7A	7.7292619	1.84E-31
SERPINB4	7.2403485	1.11E-35
TCN1	6.0285633	2.81E-25
S100A12	6.1465832	2.11E-27
SPRR2C	5.6754507	2.57E-20
S100A9	5.6818349	2.79E-26
GDA	4.6583229	5.02E-17
CXCL8	3.2690672	1.06E-13
OASL	4.2746268	4.51E-20
VNN3	3.6742317	7.73E-25
IGFL1	4.0549227	2.80E-10
KYNU	3.5853919	3.48E-22
AKR1B10	5.004838	2.69E-17
C10orf99	4.4983342	1.49E-14
LCN2	4.0258133	2.16E-16
RHCG	3.7233847	4.23E-14
CHI3L2	3.3260717	1.10E-19
DSC2	2.982965	2.04E-10
SERPINB13	3.0394025	1.21E-19
LTF	4.0099324	5.07E-12
PI3	4.7793873	2.32E-23
SERPINB3	4.7749119	7.24E-22
IL36G	4.0033896	7.42E-20
FCHSD1	2.8362518	1.05E-20
SPRR3	2.8543905	1.30E-23
KRT16	3.6926775	1.31E-16
TGM1	2.6884842	6.96E-19
INA	2.4922918	2.18E-11
SPRR2B	3.7003294	1.97E-17
KCNJ15	2.6028706	3.04E-11
CLEC7A	2.6201033	1.26E-20
FOXE1	2.6804964	3.72E-11
SLAMF7	2.4595914	1.97E-17
SLC26A9	2.3476521	5.32E-13
HPSE	3.3000869	4.44E-20
IFI27	2.7588256	1.11E-19
S100A8	2.9679482	6.93E-20
CD24	2.3216298	1.82E-14
LCE3D	3.4986925	3.55E-17
WNT5A	2.3868637	2.29E-20
MPZL2	2.1814018	8.87E-14
SLC6A14	2.384029	7.88E-12
SLC7A11	2.0531969	1.06E-09
KRT6B	1.955312	5.23E-14
GDPD3	2.0002089	8.87E-13
S100A7	2.726852	3.74E-14
GBP1	1.8934123	2.57E-14
SH3PXD2A-AS1	2.0013259	1.69E-13
SPRR1A	2.2181	7.71E-12
KRT6A	2.3307201	8.05E-10
CNFN	2.2173269	1.82E-16
IL36RN	2.0202313	1.16E-12
PLSCR1	1.7044447	2.97E-12
SPRR2G	1.73	2.96E-12

Appendix 2. Supplementary Data of Pathway Enrichment & Drug Repurposing

Table 6. Pathway Enrichment Result after Analysis of 66 DEGs.

nGenes	Pathways	Genes
17	Skin development	LCE3D SPRR3 SPRR1A SPRR2B TGM1 HPSE CNFN WNT5A SPRR2G FOXE1 KRT16 S100A7 PI3 DSC2 KRT6B SERPINB13 KRT6A
15	Keratinocyte differentiation	LCE3D SPRR3 SPRR1A SPRR2B TGM1 CNFN SPRR2G KRT16 WNT5A S100A7 PI3 DSC2 KRT6B SERPINB13 KRT6A
14	Defence response to other organism	OASL LTF LCN2 S100A12 C10orf99 KRT6A GBP1 IFI27 PLSCR1 IL36RN S100A7 CLEC7A S100A8 S100A9
12	Keratinization	TGM1 CNFN SPRR2G LCE3D SPRR3 SPRR1A KRT16 SPRR2B PI3 DSC2 KRT6B KRT6A
11	Cornification	TGM1 PI3 DSC2 SPRR2G LCE3D SPRR3 SPRR1A KRT6B KRT16 SPRR2B KRT6A
10	Antimicrobial humoral response	CXCL8 LTF S100A9 S100A12 KRT6A IL36RN S100A7 PI3 S100A8 LCN2

(nGenes denotes the number of genes out of 66DEGs present in particular pathway)

Table 7. Top 10 out of 50 potential drug molecules after CMap analysis by L1000 tool.

CMap name	Mean Connectivity Score	p-value	Specificity
Resveratrol	-0.568	0	0.0093
LY-294002	-0.287	0	0.0307
PNU-0251126	-0.393	0.00016	0
0198306-0000	-0.612	0.00052	0
Tiabendazole	-0.572	0.00101	0
Monobenzone	-0.539	0.00173	0
Parthenolide	-0.393	0.00209	0.0483
Doxycycline	-0.3	0.00332	0.0113
Methotrexate	-0.305	0.00445	0.0274
Y-27632	-0.62	0.00551	0.0055

REFERENCES

[1] R. Parisi, D. P. M. Symmons, C. E. M. Griffiths, and D. M. Ashcroft, "Global epidemiology of psoriasis: A systematic review of incidence and prevalence," *J. Invest. Dermatol.*, 2013, doi: 10.1038/jid.2012.339.

[2] Institute for Health Metrics and Evaluation (IHME), "Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2017 (GBD 2017) Results.," 2018.

[3] G. J. Nevitt and P. E. Hutchinson, "Psoriasis in the community: Prevalence, severity and patients' beliefs and attitudes towards the disease," *Br. J. Dermatol.*, 1996, doi: 10.1111/j.1365-2133.1996.tb03826.x.

[4] J. L. Harden, J. G. Krueger, and A. M. Bowcock, "The immunogenetics of Psoriasis: A comprehensive review," *Journal of Autoimmunity*. 2015, doi: 10.1016/j.jaut.2015.07.008.

[5] M. A. Lowes, C. B. Russell, D. A. Martin, J. E. Towne, and J. G. Krueger, "The IL-23/T17 pathogenic axis in psoriasis is amplified by keratinocyte responses," *Trends in Immunology*. 2013, doi: 10.1016/j.it.2012.11.005.

[6] W. H. Boehncke and M. P. Schön, "Psoriasis," *The Lancet*. 2015, doi: 10.1016/S0140-6736(14)61909-7.

[7] L. Dubertret *et al.*, "European patient perspectives on the impact of psoriasis: The EUROPSO patient membership survey," *British Journal of Dermatology*. 2006, doi: 10.1111/j.1365-2133.2006.07405.x.

[8] K. M. Wittkowski *et al.*, "Clinical symptoms of skin, nails, and joints manifest independently in patients with concomitant psoriasis and psoriatic arthritis," *PLoS One*, 2011, doi: 10.1371/journal.pone.0020279.

[9] J. F. Merola, A. Qureshi, and M. E. Husni, "Underdiagnosed and undertreated psoriasis: Nuances of treating psoriasis affecting the scalp, face, intertriginous areas, genitals, hands, feet, and nails," *Dermatologic Therapy*. 2018, doi: 10.1111/dth.12589.

[10] V. Natarajan, A. Nath, D. Thappa, R. Singh, and S. Verma, "Coexistence of onychomycosis in psoriatic nails: A descriptive study," *Indian J. Dermatol. Venereol. Leprol.*, 2010, doi: 10.4103/0378-6323.72468.

[11] T. R. Bedi, "Clinical profile of psoriasis in North India," *Indian J. Dermatol. Venereol. Leprol.*, 1995.

[12] D. Pariser, B. Schenkel, C. Carter, K. Farahi, T. M. Brown, and C. N. Ellis, "A multicenter, non-interventional study to evaluate patient-reported experiences of living with psoriasis," *J. Dermatolog. Treat.*, 2016, doi: 10.3109/09546634.2015.1044492.

[13] I. Schäfer, S. J. Rustenbach, M. Radtke, J. Augustin, G. Glaeske, and M. Augustin, "Epidemiologie der Psoriasis in Deutschland - Auswertung von Sekundärdaten einer gesetzlichen Krankenversicherung," *Gesundheitswesen*, 2011, doi: 10.1055/s-0030-1252022.

[14] K. Kubota *et al.*, "Epidemiology of psoriasis and palmoplantar pustulosis: A nationwide study using the Japanese national claims database," *BMJ Open*, 2015, doi: 10.1136/bmjopen-2014-006450.

[15] D. L. *et al.*, "European patient perspectives on the impact of psoriasis: The EUROPSO patient membership survey," *British Journal of Dermatology*. 2006.

[16] M. Moradi, F. Rencz, V. Brodszky, A. Moradi, O. Balogh, and L. Gulácsi, "Health status and quality of life in patients with psoriasis: An Iranian cross-sectional survey," *Arch. Iran. Med.*, 2015, doi: 0151803/AIM.004.

[17] M. Augustin *et al.*, "Epidemiology and Comorbidity in Children with Psoriasis and Atopic Eczema," *Dermatology*, 2015, doi: 10.1159/000381913.

[18] S. Shaharyar *et al.*, "Subclinical cardiovascular disease in plaque psoriasis: Association or causal linkα," *Atherosclerosis*. 2014, doi: 10.1016/j.atherosclerosis.2013.10.023.

[19] R. M. Robati, M. Partovi-Kia, H. R. Haghighatkhah, S. Younespour, and F. Abdollahimajd, "Increased serum leptin and resistin levels and increased carotid intima-media wall thickness in patients with psoriasis: Is psoriasis associated with atherosclerosis?," *J. Am. Acad. Dermatol.*, 2014, doi: 10.1016/j.jaad.2014.06.006.

[20] O. Ahlehoff *et al.*, "Psoriasis and risk of atrial fibrillation and ischaemic stroke: A Danish Nationwide Cohort Study," *European Heart Journal*. 2012, doi: 10.1093/eurheartj/ehr285.

[21] K. Wolk, L. Mallbris, P. Larsson, A. Rosenblad, E. Vingård, and M. Ståhle, "Excessive body weight and smoking associates with a high risk of onset of plaque psoriasis," *Acta Derm. Venereol.*, 2009, doi: 10.2340/00015555-0711.

[22] T. Tasliyurt *et al.*, "Erectile dysfunction in patients with psoriasis: Potential impact of the metabolic syndrome," *Eur. Rev. Med. Pharmacol. Sci.*, 2014.

[23] J. F. Fowler *et al.*, "The impact of psoriasis on health care costs and patient work loss," *J. Am. Acad. Dermatol.*, 2008, doi: 10.1016/j.jaad.2008.06.043.

[24] W. H. Organization, "Global report on psoriasis. 2016," *WHO Library Cataloguing-in-Publication Data*. 2016.

[25] Y. Xiao *et al.*, "A novel significance score for gene selection and ranking," *Bioinformatics*, 2014, doi: 10.1093/bioinformatics/btr671.

[26] J. Rung and A. Brazma, "Reuse of public genome-wide gene expression data," *Nature Reviews Genetics*. 2013, doi: 10.1038/nrg3394.

[27] T. Barrett *et al.*, "NCBI GEO: Archive for high-throughput functional genomic data," *Nucleic Acids Res.*, 2009, doi: 10.1093/nar/gkn764.

[28] N. V. Botchkareva, "The Molecular Revolution in Cutaneous Biology: Noncoding RNAs: New Molecular Players in Dermatology and Cutaneous Biology," *Journal of Investigative Dermatology*. 2017, doi: 10.1016/j.jid.2017.02.001.

[29] M. D. Adams *et al.*, "Complementary DNA sequencing: Expressed sequence tags and human genome project," *Science (80-.).*, 1991, doi: 10.1126/science.2047873.

[30] Q. Pan, O. Shai, L. J. Lee, B. J. Frey, and B. J. Blencowe, "Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing," *Nat.*

Genet., 2008, doi: 10.1038/ng.259.

[31] M. Sultan *et al.*, "A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome," *Science (80-.).*, 2008, doi: 10.1126/science.1160342.

[32] T. Lappalainen *et al.*, "Transcriptome and genome sequencing uncovers functional variation in humans," *Nature*, 2013, doi: 10.1038/nature12531.

[33] M. Melé *et al.*, "The human transcriptome across tissues and individuals," *Science (80-.).*, 2015, doi: 10.1126/science.aaa0355.

[34] R. Sandberg, "Entering the era of single-cell transcriptomics in biology and medicine," *Nature Methods*. 2014, doi: 10.1038/nmeth.2764.

[35] V. Romanov, S. N. Davidoff, A. R. Miles, D. W. Grainger, B. K. Gale, and B. D. Brooks, "A critical comparison of protein microarray fabrication technologies," *Analyst*, 2014, doi: 10.1039/c3an01577g.

[36] I. Barbulovic-Nad, M. Lucente, Y. Sun, M. Zhang, A. R. Wheeler, and M. Bussmann, "Bio-microarray fabrication techniques - A review," *Critical Reviews in Biotechnology*. 2006, doi: 10.1080/07388550600978358.

[37] R. Lowe, N. Shirley, M. Bleackley, S. Dolan, and T. Shafee, "Transcriptomics technologies," *PLoS Comput. Biol.*, 2017, doi: 10.1371/journal.pcbi.1005457.

[38] F. Ozsolak and P. M. Milos, "RNA sequencing: Advances, challenges and opportunities," *Nature Reviews Genetics*. 2011, doi: 10.1038/nrg2934.

[39] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: A revolutionary tool for transcriptomics," *Nature Reviews Genetics*. 2009, doi: 10.1038/nrg2484.

[40] T. Hashimshony, F. Wagner, N. Sher, and I. Yanai, "CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification," *Cell Rep.*, 2012, doi: 10.1016/j.celrep.2012.08.003.

[41] S. A. Byron, K. R. Van Keuren-Jensen, D. M. Engelthaler, J. D. Carpten, and D. W. Craig, "Translating RNA sequencing into clinical diagnostics: Opportunities and challenges," *Nature Reviews Genetics*. 2016, doi: 10.1038/nrg.2016.10.

[42] C. Mimoso, D. D. Lee, J. Zavadil, M. Tomic-Canic, and M. Blumenberg, "Analysis and meta-analysis of transcriptional profiling in human epidermis," *Methods Mol. Biol.*, 2014, doi: 10.1007/7651-2013-60.

[43] V. R. Iyer *et al.*, "The transcriptional program in the response of human fibroblasts to serum," *Science (80-.).*, 1999, doi: 10.1126/science.283.5398.83.

[44] S. A. Eming, P. Martin, and M. Tomic-Canic, "Wound repair and regeneration: Mechanisms, signaling, and translation," *Science Translational Medicine*. 2014, doi: 10.1126/scitranslmed.3009337.

[45] M. Blumenberg, "Skinomics: Past, present and future for diagnostic microarray studies in dermatology," *Expert Review of Molecular Diagnostics*. 2013, doi: 10.1586/14737159.2013.846827.

[46] S. Santoro *et al.*, "Laser capture microdissection for transcriptomic profiles in human skin biopsies," *BMC Mol. Biol.*, 2018, doi: 10.1186/s12867-018-0108-5.

[47] B. B. Shih, A. J. Nirmal, D. J. Headon, A. N. Akbar, N. A. Mabbott, and T. C. Freeman, "Derivation of marker gene signatures from human skin and their use in the interpretation of the transcriptional changes associated with dermatological disorders," *J. Pathol.*, 2017, doi: 10.1002/path.4864.

[48] W. R. Swindell, A. Johnston, J. J. Voorhees, J. T. Elder, and J. E. Gudjonsson, "Dissecting the psoriasis transcriptome: Inflammatory- and cytokine-driven gene expression in lesions from 163 patients," *BMC Genomics*, 2013, doi: 10.1186/1471-2164-14-527.

[49] J. Lamb *et al.*, "The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease," *Science (80-.).*, 2006, doi: 10.1126/science.1132939.

[50] K. Sijtsma and W. H. M. Emons, "Nonparametric statistical methods," in *International Encyclopedia of Education*, 2010.

[51] X. A. Qu, J. M. Freudenberg, P. Sanseau, and D. K. Rajpal, "Integrative clinical transcriptomics analyses for new therapeutic intervention strategies: A psoriasis case study," *Drug Discovery Today*. 2014, doi: 10.1016/j.drudis.2014.03.015.

[52] J. Kim *et al.*, "The Spectrum of Mild to Severe Psoriasis Vulgaris Is Defined by a Common Activation of IL-17 Pathway Genes, but with Key Differences in Immune Regulatory Genes," *J. Invest. Dermatol.*, 2016, doi: 10.1016/j.jid.2016.04.032.

[53] A. Johnston, A. M. Guzman, W. R. Swindell, F. Wang, S. Kang, and J. E. Gudjonsson, "Early tissue responses in psoriasis to the antitumour necrosis factor-α biologic etanercept suggest reduced interleukin-17 receptor expression and signalling," *Br. J. Dermatol.*, 2014, doi: 10.1111/bjd.12937.

[54] R. Bissonnette *et al.*, "Based on molecular profiling of gene expression, palmoplantar pustulosis and palmoplantar pustular psoriasis are highly related diseases that appear to be distinct from psoriasis vulgaris," *PLoS One*, 2016, doi: 10.1371/journal.pone.0155215.

[55] L. Ding *et al.*, "Rank-based genome-wide analysis reveals the association of Ryanodine receptor-2 gene variants with childhood asthma among human populations," *Hum. Genomics*, 2013, doi: 10.1186/1479-7364-7-16.

[56] R. Edgar, "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository," *Nucleic Acids Res.*, 2002, doi: 10.1093/nar/30.1.207.

[57] T. Barrett *et al.*, "NCBI GEO: Archive for functional genomics data sets - Update," *Nucleic Acids Res.*, 2013, doi: 10.1093/nar/gks1193.

[58] D. Sean and P. S. Meltzer, "GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor," *Bioinformatics*, 2007, doi: 10.1093/bioinformatics/btm254.

[59] G. K. Smyth, "limma: Linear Models for Microarray Data," in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, 2005.

[60] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *J. R. Stat. Soc. Ser. B*, 1995, doi: 10.1111/j.2517-6161.1995.tb02031.x.

[61] S. X. Ge, E. W. Son, and R. Yao, "iDEP: An integrated web application for differential

expression and pathway analysis of RNA-Seq data," *BMC Bioinformatics*, 2018, doi: 10.1186/s12859-018-2486-6.

[62] M. R. Hurle, L. Yang, Q. Xie, D. K. Rajpal, P. Sanseau, and P. Agarwal, "Computational drug repositioning: From data to therapeutics," *Clinical Pharmacology and Therapeutics*. 2013, doi: 10.1038/clpt.2013.1.

[63] F. S. Collins, "Reengineering translational Science: The time is right," *Science Translational Medicine*. 2011, doi: 10.1126/scitranslmed.3002747.

[64] J. Brent Richards, H. F. Zheng, and T. D. Spector, "Genetics of osteoporosis from genome-wide association studies: Advances and challenges," *Nature Reviews Genetics*. 2012, doi: 10.1038/nrg3228.

[65] W. Yu, M. Gwinn, M. Clyne, A. Yesupriya, and M. J. Khoury, "A navigator for human genome epidemiology [2]," *Nature Genetics*. 2008, doi: 10.1038/ng0208-124.

[66] P. Sanseau *et al.*, "Use of genome-wide association studies for drug repositioning," *Nature Biotechnology*. 2012, doi: 10.1038/nbt.2151.

[67] K. D. Cooper *et al.*, "Interleukin-1 in human skin: Dysregulation in psoriasis," *J. Invest. Dermatol.*, 1990, doi: 10.1111/1523-1747.ep12505698.

[68] T. Lane and H. J. Lachmann, "The emerging role of interleukin-1ß in autoinflammatory diseases," *Curr. Allergy Asthma Rep.*, 2011, doi: 10.1007/s11882-011-0207-6.

[69] C. Buerger *et al.*, "Interleukin-1B interferes with epidermal homeostasis through induction of insulin resistance: Implications for psoriasis pathogenesis," *J. Invest. Dermatol.*, 2012, doi: 10.1038/jid.2012.123.

[70] J. Sun, Y. Zhao, and J. Hu, "Curcumin Inhibits Imiquimod-Induced Psoriasis-Like Inflammation by Inhibiting IL-1beta and IL-6 Production in Mice," *PLoS One*, 2013, doi: 10.1371/journal.pone.0067078.

[71] A. Wojas-Pelc and J. Marcinkiewicz, "What is a role of haeme oxygenase-1 in psoriasis? Current concepts of pathogenesis," *International Journal of Experimental Pathology*. 2007, doi: 10.1111/j.1365-2613.2006.00505.x.

[72] K. J. Trouba, H. K. Hamadeh, R. P. Amin, and D. R. Germolec, "Oxidative stress and its role in skin disease," *Antioxidants and Redox Signaling*. 2002, doi: 10.1089/15230860260220175.

[73] R. P. Nair *et al.*, "Genome-wide scan reveals association of psoriasis with IL-23 and NF-κB pathways," *Nat. Genet.*, 2009, doi: 10.1038/ng.311.

[74] A. Hrenn, T. Steinbrecher, A. Labahn, J. Schwager, C. M. Schempp, and I. Merfort, "Plant phenolics inhibit neutrophil elastase," *Planta Med.*, 2006, doi: 10.1055/s-2006-946700.

LIST OF PUBLICATIONS

CONFERENCE PAPERS

1. **Pawan Singh Gangwar** and Yasha Hasija. **Big Data Analytics in Health Informatics for Precision Medicine.** Fourth International Conference on Information and Communication Technology for Intelligent Systems (ICTIS-2020), May 15th-16th, 2020 (Springer). (Accepted)
2. **Pawan Singh Gangwar** and Yasha Hasija. **Big Data Analytics and Internet of Things in Health Informatics.** Third International Conference on Soft Computing and Signal Processing (ICSCSP-2020), August 21st-22nd, 2020 (Springer). (Communicated)

BOOK CHAPTERS

1. **Pawan Singh Gangwar** and Yasha Hasija. **Deep Learning for Analysis of Electronic Health Records (EHR).** In: Deep Learning Techniques for Biomedical and Health Informatics. Studies in Big Data, vol 68. Springer 2020. (Published)
2. **Pawan Singh Gangwar** and Yasha Hasija. **Biomedical Big Data Analytics using IoT in Health Informatics.** In: Big Data Analytics for Internet of Things. Wiley 2020. (Communicated)