REVIEW ARTICLE

ISSN: 2455-8095
doi: 10.21477/ijapsr.4.1.2

Functional Genomics–Linking Genotype with Phenotype on Genome-wide Scale

Nida Tabassum Khan[1], Namra Jameel[2], Maham Jamil Khan[3]

[1]Department of Biotechnology, Faculty of Life Sciences and Informatics, Balochistan University of Information Technology
[2]Engineering and Management Sciences,(BUITEMS),Quetta, Pakistan.
OGDCL (Oil and Gas Development Company Limited)-Islamabad-Pakistan
[3]Sukkur IBA University-Undergraduate student-Sukkur-Pakistan

ABSTRACT

Functional genomics manipulates genomic data to study genes and its expression on a genome wide scale involving high-throughput methods. The keyobjective of Functional genomics is to exploit the data acquired from transcriptomic and genomic studies to explain the functions and interfaces of a genome and its corresponding phenotype.

Key Words: Genomics; Epigenomics; Proteomics; Interactomics.

ABBREVIATIONS

DNA:Deoxyribonucleic acid; NGS: Next generation sequencing; ENCODE: Encyclopedia of DNA Elements; SNP: single nucleotide polymorphism; GTG banding: Giemsa *banding;* aCGH: microarray-based comparative genomic hybridization; FISH:fluorescent *in situ* hybridization; MLPA: Multiplex ligation-dependent probe amplification; PCR: Polymerase chain reaction; dNTPs: deoxyribonucleoside 5′-triphosphates; NCBI: National Center for Biotechnology Information; *OMIM : Online Mendelian Inheritance in Man*; *dbSNP* : Database for Single Nucleotide Polymorphism; CpG: Cytosine followed by guanine; MDRE : methylation-dependent restriction enzymes; ChIP :chromatin immunoprecipitation ; RNA :Ribonucleic acid; TF: transcription factors ; LCR: locus control region; TSSs: transcription start sites; TFBSs: transcription factors binding sites; GFP: green fluorescent protein ; mRNA: messenger RNA; cDNA: Complementary deoxyribonucleic acid; SAGE: serial analysis of gene expression; qPCR : quantitative real-time PCR; ELISA: enzyme-linked immunosorbent assay; 2- DE: two-dimensional gel electrophoresis; CT method: cycle thresholdmethod ; MS : mass spectrometry; *m*/z: mass-to-charge; MSIA: mass spectrometric immunoassay; Y2H: two-hybrid system; HTS: High-throughput screening; CO2: Carbondioxide; CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats.

INTRODUCTION

Functional genomics glances at the functional aspects of the genes and their expression to explain transcription, translation, protein complex interactions etc., based on genomic information (Hieter and Boguski, 1997). Research in Functional genomics benefits us to understand DNA function and its interaction (White, 2001). A genome-wide methodology is adopted to study the role of number of genes rather than concentrating on a single gene as previously done (Colebatch *et al*., 2002). Functional genomics holds the potentials to reveal that an organism's genome influences its biological function (Kao, 1999). This could serve as a viable tool in the future for the treatment of human genetic diseases (Dean, 2001). Functional genomics explores mechanisms of gene expression and how their expression levels differs with respect to cell types and states, functional roles of different genes ,gene regulation, interaction of genes and its products etc., (Fiehn, 2002; Nielsen and Olsson, 2002).

TECHNOLOGIES IN FUNCTIONAL GENOMICS

To comprehend the information stored in DNA, numerous high-throughput methods are used in Functional genomics for genome analysis (Morozova and Marra, 2008). Furthermore development in the field of bioinformatics

Corresponding author
Nida Tabassum Khan, Phone: 03368164903
Email: nidatabassumkhan@yahoo.com

Received: 12-10-2018 **Accepted:** 25-11-2018 **Available Online:** 01-01-2019

provides a precise and comprehensive functional analysis in genomics, epigenomics, proteomics, and interactomics (Hawkins *et al*., 2010). This is crucial for filling the gaps in the knowledge about dynamic biological processes at both cellular and organismal level (Ohashi *et al*.,2015). Although human genome is about 99.9% identical, the remaining 0.1% is the reason of difference between people caused by different variants (Collins and McKusick, 2001). Since 2003, the complete sequence of human genome, its annotation and increased advancement of sequencing technologies (i. e., Sanger and Next-generation- sequencing; NGS) have provided all the necessary conditions for the identification of all variants in human coding and non-coding sequence (Zhang *et al*., 2011). Although the technique for variant detection is now becoming a routine, the key question throughout many years concerns the function of detected variants (Metzker, 2010). The resource of important information about functional genomics are several large-scale projects, for instance, the ENCODE project, the main goal of which was to identify all the efficient elements, including regulatory elements in both coding and non-coding regions (ENCODE Project Consortium, 2004). According to another, the 1000 Genomes Project, there are about 20,000–23,000 variants in synonymous and nonsynonymous regions of the human genome (1000 Genomes Project Consortium, 2012). Even though not all of them are functionally meaningful, 530–610 of the variants have functional impact by causing inframe deletions and insertions, premature stop codons, frameshifts, or by disrupting splice sites (Stephenson, 2008). Despite numerous studies, scientists are still facing a huge challenge in unravelling what the sequence means and in deciding whether or not a found variant is pathogenic, a pathogenic variant can lead to disease or cause a number of disorders (Cooper and Shendure, 2011). However, understanding of pathogenic mechanisms creates an opportunity to prevent severe consequences by developing novel diagnostic tools and by designing highly effective treatments for the disease (Maurano *et al*., 2012).

FROM VARIANT DETECTION TO FUNCTIONAL GENOME ANALYSIS

Variants in exonic regions and intronic genome sequences range from single nucleotide changes to large, microscopically visible, chromosomal aberration (Hindorff *et al*., 2009). These variants may influence the biological function of a gene. They can be either beneficial (e.g., single nucleotide polymorphism; SNP) with no negative effect on the phenotype, or pathogenic (e.g., nonsense variant) – resulting in a number of different disorders and diseases (Liu, 2007). Depending on the variant type and locus, there are numerous different genetic methods and tools for the variant detection (Long and Langley, 1999). For example, due to its simplicity the most frequent method for the analysis of a large (>5 Mb) chromosomal aberration is karyotype analysis by using the GTG banding technique (Speicher *et al*.,1996) .Other molecular genetic methods, such as microarray-based comparative genomic hybridization (aCGH) (Theisen, 2008) or fluorescent *in situ* hybridization (FISH) (O'connor, 2008), should be applied for a more accurate analysis. However, these methods have some significant limitations: the aCGH does not detect mosaicism, balanced translocations and inversions, while the FISH requires specific probes (Oostlander *et al*., 2004). Moreover, for particular variant detection another molecular genetic methods might be applicable, which include restriction enzyme assay, Multiplex ligation dependent probe amplification (MLPA), even though many of the tests are based on the Polymerase chain reaction (PCR) and its variants (e.g., multiplex PCR) (Iafrate *et al*., 2004). Although researchers can easily plan their assay in the case of particular variants, they are facing some challenges in the study of unspecified variants (Grompe, 1993). Sequencing is considered as a "gold standard" method for the documentation of known as well as unspecified variants in the genomic DNA (Lappalainen *et al*., 2013). In accordance with the previous statement, the Sanger or Next-Generation Sequencing (NGS) techniques can be used (Hawkins *et al*., 2010). The concept behind these two methods is similar. During the polymerase chain reaction, which consists of several cycles of sequential DNA replication, DNA polymerase catalyzes the complementary incorporation of fluorescently-labeled deoxyribonucleoside 5′-triphosphates (dNTPs) into the DNA template. For each cycle, a colour of the labeled DNA fragment is recorded by a detector, thus determining nucleotide in the sequence (Van *et al*., 2014). The main difference between the conventional (i.e., Sanger) technology and the NGS is that the latter is not limited to a single DNA fragment but analyzes millions of fragments in massively parallel sequencing technology (Wheeler *et al*., 2008). These two sequencing methods are widely used all over the world. Even so, it is considered that in a small-scale project it is more eligible to use the Sanger sequencing system because of its accuracy. On the other hand, in large-scale projects this research method would be expensive and time-consuming, therefore the NGS needs to be applied (Schuster, 2007). General progress in technology achieved in some strategies of the next-generation DNA sequencing has a huge impact on genetic research. Recently, the most widely used platforms have been Roche/454 Life Science (Liu *et al*., 2012), Applied Biosystems SOLiD (Pandey *et al*., 2008), and Illumina Genome Analyzer (Quail *et al*., 2009). Another DNA

sequencing technology has been lately developed by Ion Torrent (Poptsova *et al*., 2014). Nevertheless, "sequencing-by-synthesis" used by Illumina currently is one of the most popular NGS platform (Cronn *et al*., 2008). First of all, a randomly fragmented DNA is ligated with specific adaptors and amplified by the use of PCR. Secondly, a preformed DNA library should be immobilized on the beads or arrays, thus generating clusters of identical DNA fragments. These clusters are then read by sequential cycles of nucleotide incorporation, washing, and detection, where the number of cycles eventually determines the read length (Quail *et al*., 2008). In order to understand the genome structure, function, or evolution, it is not enough to obtain the DNA sequencing data through the NGS: but there is also a need for deep and precise analysis using bioinformatics approaches. The key path to successful sequence analysis is to align the sequence of interest with another sequence whose function is known (usually termed as the reference genome) (Howe *et al*., 2013). It might be very useful when the gene function is unknown but is evolutionary related to another gene whose function is defined (Darling, 2010). In such a case, it can be suspected that the unknown gene has the same or similar function. Furthermore, the sequences might be scanned in order to find the significant matches between the components of a sequence that have been previously described as having a huge impact on the genomics function (Delcher *et al*.,1999). In order to compare the data, it is necessary to search for information in different biomedical databases. One of the biggest sources of biomedical and genomic information is the NCBI (National Center for Biotechnology Information), which provides access to numerous databases such as PubMed, Entrez Gene, OMIM, Variation Viewer, dbSNP, and others (Wheeler *et al*.,2003).

EPIGENOMICS

For functional analysis, epigenetic modifications such as DNA methylation and histone modifications is taken into account, because they affect gene expression without any changes in the underlying DNA sequence (Bjornsson *et al*., 2014). DNA methylation, which usually occurs in the context of densely situated CpG dinucleotide (i. e., CpG islands), correlates with transcriptional suppression (Fuks, 2005). In order to detect DNA methylation status, unmethylated cytosines are converted into uracil by using sodium bisulfite, because methylated cytosine is resistant to this impact. Additionally, methylation-dependent restriction enzymes (MDRE) are highly effective for DNA methylation analysis. These enzymes, e.g., *HpaII* and *MspI*, recognize and simply digest the methylated DNA (Pu and Clark, 2003). Usually, MDRE or even more frequently used bisulfite conversion is the first step for many subsequent methods such as methylation-specific PCR, sequencing, bead array, etc (Laird, 2010). Histone modifications by acetylation, phosphorylation, methylation, ubiquitination, and others – are another cause of epigenetically-regulated genes (Peterson and Laniel, 2004). Depending on the modification type and locus, gene expression can be either be activated or repressed (Berger, 2002). The most common method for the investigation of histone modifications is chromatin immunoprecipitation (ChIP) based on the interaction between the antigen (of associated with DNA target protein) and the antibody (specific to target modified protein) (Nelson *et al*., 2006). After the precipitation, the genomic DNA is released for further research hinged on microarray analysis (ChIPchip), sequencing (ChIP-seq), or quantitative PCR. Although these methods are high-throughput, the dependence on a specific antibody sometimes limits the use of the ChIP (Collas and Dahl, 2008).

TRANSCRIPTOMICS

When the human genome was fully sequenced, the focus of attention shifted towards identifying and annotating its functional DNA elements, including those that regulate gene expression (Schnappinger, 2008). Identification of such elements is a vitally important step towards elucidating pathogenic pathways that affect human health (Alderton, 2010). All the RNA-level processes, including transcription activation or inhibition, mRNA processing and its transport are regulated by different functional elements of the genomic DNA (Wang *et al*., 2009). Nevertheless, the highest regulation occurs at the transcriptional initiation level through several regulative elements, which are called the *cis*-acting regulatory sequence and *trans*-factors (EpSTEIN *et al*., 1968). *Trans*-factors such as transcription factors (TF), activators, and repressors (including co-activators and co-repressors) interact with specific DNA regions, i.e., *cis*-acting regulatory sequence that includes core promoter (with a TATA box and other binding elements), proximal promoter, enhancer, silencer, insulator, and locus control region (LCR) (Maston *et al*., 2006). Investigation of these regulatory elements may be a challenge for the scientists because of the difficulties in identifying the position of transcription start sites (TSSs) and transcription factors binding sites (TFBSs) in the core promoter (Heintzman *et al*., 2007). However, there are several experimental and bioinformatical approaches. First of all, a comparative bioinformatical approach is necessary for the study of the regulatory elements. This type of research is usually based on constructing alignments between orthologous sequences because sequence homology provides valuable evidences to gene function

analysis (Lenhard *et al*., 2003). Nevertheless, a deeper understanding of regulatory elements requires laboratory investigations. It is believed that every TFBS could be detected by the above-mentioned ChIP method (Hu *et al*., 2010). Theoretically, depending on immunoprecipitation of the target protein, the core promoters, enhancers, silencers, insulators and LCRs could be determined (Valouev *et al*., 2008). Furthermore, epigenetic markers can be helpful in detecting TSSs in the core promoter and enhancer loci, because TSSs of actively transcribed genes are marked by H3K4me3 and H3K27ac, while enhancers by H3K4me1 and H3K27ac (Hawkins *et al*., 2010). Another very frequent functional assay of the regulatory element is based on the transgenesis of a specific reporter- gene (e. g., the gene of the green fluorescent protein – GFP or luciferase) into the target regulatory sequence (Aparicio *et al*., 1995). After the translation, activity of the reporter-gene is measured, e.g., by fluorescence of the GFP, with the purpose to determine if the examined region contains elements that alter reporter-gene expression (Rosenthal, 1987). Substantial information about functional genomics can be obtained through the analysis of the messenger RNA (mRNA) or cDNA, which is copied from the mRNA by reverse transcription PCR (Wong and Medrano, 2005). Therefore researchers often choose to test the mRNA or cDNA rather than DNA, because RNA analysis may be more eligible for a gene that has many small exons and it can also reveal abnormal splicing (Fraser *et al*., 2000). For many years there have been some standard methods for measuring the mRNA expression: Northern blotting (Brown *et al*., 2004), serial analysis of gene expression (SAGE) (Velculescu *et al*., 1995) as well as quantitative real-time PCR (qPCR) (Heid *et al*., 1996) among them. The SAGE method is based on the conversion of an RNA molecule into a short unique tag, while Northern blotting – on hybridization with radioactive probe. This allows to perform quantitative analysis by counting the number of tags and measuring intensity of band, respectively (Kozian and Kirschbaum, 1999). However, both these methods are characterized as low-throughput. Nevertheless, for the mRNA quantitation and gene expression evaluation the "gold standard" is qPCR, which is fast, very sensitive, and highly reproducible (Mestdagh *et al*., 2009). The principle of this method is that during the reverse transcriptional reaction, complementary single-stranded cDNA from the RNA template is synthesized. The cDNA is necessary for subsequent use in quantitative PCR (Taylor *et al*., 2010). The aim of this reaction is to measure fluorescence intensity that is directly proportional to the amount of cDNA in the sample (Mestdagh *et al*., 2008). There are two strategies for qPCR data analysis: absolute quantification (based on the calibration curve) and relative quantification (based on the comparison with reference sample) (Pfaffl, 2012). For the relative gene expression level calculation, the most convenient way is comparative *CT* method. This method relies on comparing the CT values of the target and reference samples, using a reference (endogenous housekeeping) gene as the normalizer. Finally, the method results in the fold change of target gene expression relative to a reference sample, normalized to a housekeeping gene (Schefe *et al*., 2006). Acceleration of high-throughput technologies such as cDNA microarray and RNA sequencing (RNA-seq), which also provides the possibility of transcriptional characterization, very often replaces preceding methods (Rabbani *et al*., 2003). Results obtained by a cDNA microarray assay provide important genome- wide information about the changes of gene expression in various cell lines and in different stages of development (DeRisi *et al*., 1996). This method is based on hybridization of fluorescently labelled cDNA with the particular oligonucleotides (probe) on the specific microarray. The amount of hybridization recorded for a specific probe is proportional to the number of DNA fragments in the sample. In this way, the obtained absolute hybridization values give an opportunity to detect genetic variation in the human genome (Schena *et al*., 1995). Despite the great advantages of cDNA microarray, high-throughput RNA sequencing based on different NGS systems is also increasingly used (Wilhelm and Landry, 2009). The RNA sequencing results in a number of short reads. Aligned to a reference genome, they produce a specific transcription map that corresponds to the transcriptional structure and gene expression level (Ozsolak *et al*., 2009). It means that this technique is appropriate for gene, transcripts (including alternative gene spliced transcripts), or allele-specific expression identification. Moreover, it is possible to accurately measure translation of transcripts (Wold and Myers, 2007). As each method has both advantages and disadvantages, the last one is not an exception. The problems in RNA-seq are often related with high sequence similarity between alternative spliced isoforms or difficulties in data analysis (Wall *et al*., 2009).

PROTEOMICS AND INTERACTOMICS

From the functional point of view, analysis of proteomics and interactomics is as vitally important as previously described analysis of genomics, epigenomics, and transcriptomics, because some studies show that gene expression at DNA or mRNA levels is substantially unchanged, although it affects the protein function and vice versa (Ivanov *et al*., 2011). Proteins perform a vast array of functions within organisms, though abnormal protein expression that occurs due to post-transcriptional modifications or protein interaction with another protein or

nucleic acids disrupts cell function (Filipowicz *et al*., 2008). Depending on the intent of the experiment, there are two well-known strategies for protein quantification: immunoassays or antibody-free detection methods (Schuurs, 1997). Immunoassay, such as the enzyme-linked immunosorbent assay (ELISA), is a widely-used method due to its high sensitivity and strong specificity (Reen, 1994). However, sometimes researchers can face the problem when no antibody exists for the protein of interest. In such cases the solution is antibody-free methods (Shi *et al*., 2012). Firstly, compared to one-dimensional protein separation method, two-dimensional gel electrophoresis (2- DE), which separates protein by two properties in 2D gels is more effective (Gygi *et al*., 2000). However, the most common and comprehensive analytical tool for protein detection, identification, and quantification is mass spectrometry (MS) that measures mass-to-charge (*m*/*z*) ratio of ions (Ong and Mann, 2005). Advancement of MS gives an opportunity to achieve a greater throughput of samples with high accuracy and precision. Additionally, it is considered that MS methodology is rapid and reliable for large-scale studies. Furthermore, due to its advantages MS is very often combined with another technique (Aebersold and Mann, 2003). For instance, some studies consist of antibody-based purification and mass spectrometry analysis termed mass spectrometric immunoassay (MSIA) (Domon and Aebersold, 2006; Anderson *et al*., 2004). An important step towards characterization of the protein function is the identification of the protein interaction network consisting of different proteins (Stelzl *et al*., 2005). The most frequent system for detection of interacting proteins in living yeast cells is the two-hybrid system (Y2H) (Bartel and Fields,1997). The aim of such investigation is to create genetically modified yeast strains on a selective medium. In such a system, two interacting proteins bound to specific domains switch on polymerase II, which subsequently activate the transcription of a reporter gene, whose transcription leads to a specific phenotype (e.g., changed color) (Yang and Fields,1995). Furthermore, proteins interact also with nucleic acids; (DNA and RNA). In functional approach, the most important interactions are between DNA and transcription factors or regulatory elements (Johnson *et al*., 2007). In the case of RNA, it is necessary to test interactions between this nucleic acid and ribosome, or other RNA binding proteins (Dassi, 2016). The analysis of both DNA-protein and RNA-protein interactions is based on similar techniques. Previously mentioned high-throughput immunoprecipitation of the nucleic acid and protein complex is increasingly becoming the method of choice for the detection of TFBSs and histone modification (O'Neill and Turner, 2003). Subsequent microarray, or NGS analysis, enables the identification of a particular locus, i. e., the region that is specifically interacting with the protein of interest (Wu *et al*., 2014). However, the main limitation of the ChIP method is the dependence on antibody specificity (Poetz *et al*., 2005). Above mentioned functional genomics techniques could be useful; along with the completion of Human Genome Project, large quantity of data about the genetic basis of human have been acquired (Watson, 1990). These studies produce massive amounts of data, e.g. expression values from tens of microarray chips, each with thousands of probes, or tens of millions of very short sequence reads from HTS machines (Langenberger *et al*., 2010). These data can only help us gain insight into underlying biological processes, if they are carefully recorded and stored in databases, along with the experimental workflows employed and annotated detail for each sample. This will then allow the data to be queried, compared, analyzed, interpreted and shared by the research community (Aggarwal and Lee, 2003).

APPLICATIONS OF FUNCTIONAL GENOMICS IN LIFE SCIENCES

Plant function, development and regulation could be easily understood with the help of Functional genomics (Fiehn *et al*., 2000).

Functional genomics tools helps in recognizing useful metabolic pathways and modify them in specific plants or crops to enhance food quality (Slade and Knauf, 2005).

Functional genomics and RNA sequencing could be used to facilitate the breeding process replacing the traditional marker assisted methods with cost effective whole genome or transcriptome studies (Andersen and Lübberstedt, 2003).

Genomes/transcriptome sequencing of different plants and algae, helps in discovering new sources for biofuels (Mukhopadhyay *et al*., 2008).

Sequencing projects of individual plants genomes/ transcriptomes from a specific ecosystem helps in determining the genetic variability of the system (Feder and Mitchell, 2003).

Genome/transcriptome analysis of different species helps to trace particular genes and pathways to identify novel interactions (Ungerer *et al*., 2008).

The application of NextGen sequencing helps in comprehending organism's taxonomy and its evolution (Altenhoff and Dessimoz, 2009).

Functional genomics identifies specific pathways and genes for novel drug production under controlled conditions (Evans and Relling, 1999).

Mitigating climate change, especially from the view point of increasing CO2 consumption by plants and algae could be achieved by modifying existing plant and algal systems (Zhao and Poh, 2008).

De novo design of organisms at designing biological systems and components that do not exist in nature. Such design of organisms is limited to viruses, mycoplasma and perhaps in the near future bacteria (Keasling, 2010).

CONCLUSION

It is believed that successful functional genome analysis discovers genetic basic for human health by filling the gaps in knowledge about pathogenic pathways between genes, proteins, and their interaction network. There are a lot of different methods and tools for accurate functional analysis. Despite huge analytical progress, these methods have certain limitations. Thus, in order to extend the limits of current techniques, some high-throughput technologies such as quantitative real-time polymerase chain reaction, next-generation sequencing or mass spectrometry have been developed, which provide an opportunity to perform genome-wide functional analysis. Furthermore, model systems such as CRISPR-Cas9 or animal models are required for an extensive functional interpretation of genome sequence variants. However, in processing large amounts of data researchers are still facing the problem, that usually is very complicated and time consuming. For this reason, there is a need of continuous improvement in technology and development of more efficient analytical tools. It should be noted that for more comprehensive results it is essential to use complex methodologies that complement each other's shortcomings.

REFERENCES

1. 1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature, 491(7422), 56-65.
2. Aebersold R, & Mann M (2003). Mass spectrometry-based proteomics. *Nature*, 422(6928):198-207.
3. Aggarwal K & Lee HK (2003). Functional genomics and proteomics as a foundation for systems biology. *Briefings in Functional Genomics*. 2(3):175-184.
4. Alderton GK (2010). Transcriptomics: common disease pathogenesis pathways. *Nature Reviews Cancer*. 10(6):387-387.
5. Altenhoff AM & Dessimoz C (2009). Phylogenetic and functional assessment of orthologs inference projects and methods. *PLoS computational biology*. 5(1):e1000262.
6. Andersen JR & Lübberstedt T (2003). Functional markers in plants. *Trends in plant science*. 8(11):554-560.
7. Anderson NL, Anderson NG, Haines LR, Hardie DB, Olafson RW, & Pearson TW (2004). Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *Journal of proteome research*, 3(2):235-244.
8. Aparicio S, Morrison A, Gould A, Gilthorpe J, Chaudhuri C, Rigby P, Brenner S (1995). Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, Fugu rubripes. *Proceedings of the National Academy of Sciences*. 92(5):1684-1688.
9. Bartel P & Fields S (Eds.) (1997). The yeast two-hybrid system. *Oxford University Press*, USA.
10. Berger SL (2002). Histone modifications in transcriptional regulation. *Current opinion in genetics & development*. 12(2):142-148.
11. Bjornsson HT, Fallin MD & Feinberg AP (2004). An integrated epigenetic and genetic approach to common human disease. *TRENDS in Genetics*, 20(8):350-358.
12. Brown T, Mackey K & Du T (2004). Analysis of RNA by Northern and slot blot hybridization. *Current protocols in molecular biology*. 4-9.
13. Colebatch G, Trevaskis B & Udvardi M (2002). Functional genomics: tools of the trade. *New Phytologist*, 153(1):27-36.
14. Collas P & Dahl JA (2008). Chop it, ChIP it, check it: the current status of chromatin immunoprecipitation. *Front Biosci*. 13(17):929-943.
15. Collins FS & McKusick VA (2001). Implications of the Human Genome Project for medical science. *Jama*. 285(5):540-544.
16. Cooper GM & Shendure J (2011). Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nature Reviews Genetics*, 12(9):628-640.
17. Cronn R, Liston A, Parks M, Gernandt DS, Shen R & Mockler T (2008). Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic acids research*, 36(19):e122-e122.
18. Darling AE, Mau B & Perna NT (2010). ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS one*. 5(6):e11147.
19. Dassi E (Ed.) (2016). Post-transcriptional gene regulation. *Humana Press*.
20. Dean NM (2001). Functional genomics and target validation approaches using antisense oligonucleotide technology. *Current opinion in biotechnology*, 12(6):622-625.
21. Delcher A., Kasif S, Fleischmann RD, Peterson J, White O & Salzberg SL (1999). Alignment of whole genomes. *Nucleic acids research*. 27(11):2369-2376.
22. DeRisi J, Penland L, Bittner M L, Meltzer PS, Ray M, Chen Y & Trent JM (1996). Use of a cDNA microarray to analyse gene expression. *Nat. genet*. 14:457-460.
23. Domon B & Aebersold R (2006). Mass spectrometry and protein analysis. *Science*. 312(5771);212-217.
24. ENCODE Project Consortium (2004). The ENCODE (ENCyclopedia of DNA elements) project. *Science*. 306(5696):636-640.
25. EpSTEIN, W. O. L. F. G. A. N. G, & Beckwith JR (1968). Regulation of gene expression. *Annual review of biochemistry*, 37(1), 411-436.
26. Evans WE, & Relling MV (1999). Pharmacogenomics: translating functional genomics into rational therapeutics. *Science*, 286(5439):487-491.
27. Feder ME & Mitchell-Olds T (2003). Evolutionary and ecological functional genomics. *Nature Reviews Genetics*. 4(8);649-655.

28. Fiehn O (2002). Metabolomics—the link between genotypes and phenotypes. In Functional Genomics (pp. 155-171). *Springer Netherlands*.
29. Fiehn O, Kopka , Dörmann P, Altmann T, Trethewey RN & Willmitzer L (2000). Metabolite profiling for plant functional genomics. *Nature biotechnology*. 18(11):1157-1161.
30. Filipowicz W, Bhattacharyya SN & Sonenberg N (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?. *Nature reviews genetics*. 9(2):102-114.
31. Fraser AG, Kamath RS, Zipperlen P, Martinez-Campos, M, Sohrmann M, & Ahringer J (2000). Functional genomic analysis of C. elegans chromosome I by systematic RNA interference. *Nature*, 408(6810):325-330.
32. Fuks, F. (2005). DNA methylation and histone modifications: teaming up to silence genes. Current opinion in genetics & development, 15(5), 490-495.
33. Grompe M (1993). The rapid detection of unknown mutations in nucleic acids. *Nature genetics*. 5(2):111-117.
34. Gygi SP, Corthals GL, Zhang Y, Rochon Y & Aebersold R (2000). Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proceedings of the National Academy of Sciences*. 97(17):9390-9395.
35. Hawkins RD, Hon GC & Ren B. (2010). Next-generation genomics: an integrative approach. *Nature Reviews Genetics*. 11(7):476-486.
36. Heid CA, Stevens J, Livak KJ & Williams PM (1996). Real time quantitative PCR. *Genome research*. 6(10), 986-994.
37. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD & Wang W (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics*. 39(3):311-318.
38. Hieter P & Boguski M (1997). Functional genomics: it's all how you read it. *Science*. 278(5338):601-602.
39. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS & Manolio TA. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences*. 106(23):9362-9367.
40. Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M & McLaren S (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, 496(7446):498-503.
41. Hu M, Yu J, Taylor JM, Chinnaiyan AM & Qin ZS (2010). On the detection and refinement of transcription factor binding sites using ChIP-Seq data. *Nucleic acids research*, 38(7):2154-2167.
42. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y & Lee C (2004). Detection of large-scale variation in the human genome. *Nature genetics*. 36(9):949-951.
43. Ivanov AS, Zgoda VG & Archakov AI (2011). Technologies of protein interactomics: a review. *Russian journal of bioorganic chemistry*, 37(1):4-16.
44. Johnson DS, Mortazavi A, Myers RM & Wold B (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science*. 316(5830):1497-1502.
45. Kao CM (1999). Functional genomic technologies: creating new paradigms for fundamental and applied biology. *Biotechnology progress*. 15(3):304-311.
46. Keasling JD (2010). Manufacturing molecules through metabolic engineering. *Science*. 330(6009):1355-1358.
47. Kozian DH & Kirschbaum BJ (1999). Comparative gene-expression analysis. *Trends in biotechnology*. 17(2):73-78.
48. Laird PW (2010). Principles and challenges of genome-wide DNA methylation analysis. *Nature Reviews Genetics*. 11(3):191-203.
49. Langenberger D, Bermudez-Santana CI, Stadler PF & Hoffmann S (2010). Identification and classification of small RNAs in transcriptome sequence data. *In Biocomputing*. 2010:80-87.
50. Lappalainen T, Sammeth M, Friedländer MR, AC't Hoen P, Monlong J, Rivas MA & Barann M. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 501(7468):506-511.
51. Lenhard B, Sandelin A, Mendoza L, Engström P, Jareborg N & Wasserman WW (2003). Identification of conserved regulatory elements by comparative genome analysis. *Journal of biology*.2(2):13.
52. Liu L, Li Y, Li S, Hu N, He Y, Pong R & Law M (2012). Comparison of next-generation sequencing systems. *BioMed Research International*. 2012.
53. Liu Z (2007). Single nucleotide polymorphism (SNP). In Aquaculture Genome Technologies (pp. 59-72). Blackwell USA.
54. Long AD & Langley CH (1999). The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. Genome Research. 9(8):720-731.
55. Maston GA, Evans SK & Green MR (2006). Transcriptional regulatory elements in the human genome. *Annu. Rev. Genomics Hum. Genet*. 7:29-59.
56. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H & Shafer A. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science*, 337(6099):1190-1195.
57. Mestdagh P, Feys T, Bernard N, Guenther S, Chen C, Speleman F & Vandesompele J (2008). High-throughput stem-loop RT-qPCR miRNA expression profiling using minute amounts of input RNA. *Nucleic acids research*, 36(21):e143-e143.
58. Mestdagh P, Van Vlierberghe P, De Weer A, Muth D, Westermann F, Speleman F & Vandesompele J (2009). A novel and universal method for microRNA RT-qPCR data normalization. *Genome biology*. 10(6):R64.
59. Metzker ML (2010). Sequencing technologies—the next generation. *Nature reviews genetics*, 11(1):31-46.
60. Morozova O & Marra MA (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics*. 92(5):255-264.
61. Mukhopadhyay A, Redding AM, Rutherford BJ & Keasling JD (2008). Importance of systems biology in engineering microbes for biofuel production. Current opinion in biotechnology. 19(3):228-234.
62. Nelson JD, Denisenko O & Bomsztyk K (2006). Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nature protocols*, 1(1):179.
63. Nielsen J & Olsson L (2002). An expanded role for microbial physiology in metabolic engineering and functional genomics: moving towards systems biology. *FEMS yeast research*. 2(2):175-181.

64. O'Neill LP & Turner BM (2003). Immunoprecipitation of native chromatin: *NChIP. Methods*, 31(1):76-82.
65. O'connor C (2008). Fluorescence in situ hybridization (FISH). *Nature Education*. 1(1):171.
66. Ohashi H, Hasegawa M, Wakimoto K & Miyamoto-Sato E (2015). Next-generation technologies for multiomics approaches including interactome sequencing. *BioMed research international*. 2015.
67. Ong SE & Mann M (2005). Mass spectrometry–based proteomics turns quantitative. *Nature chemical biology*, 1(5):252-262.
68. Oostlander AE, Meijer GA & Ylstra B (2004). Microarray-based comparative genomic hybridization and its applications in human genetics. *Clinical genetics*. 66(6):488-495.
69. Ozsolak F, Platt AR, Jones DR, Reifenberger JG, Sass LE, McInerney P & Milos PM. (2009). Direct RNA sequencing. *Nature*. 461(7265):814-818.
70. Pandey V, Nutter RC & Prediger E (2008). Applied biosystems solid™ system: ligation-based sequencing. Next Generation Genome Sequencing: Towards Personalized Medicine. 29-42.
71. Peterson CL & Laniel MA (2004). Histones and histone modifications. Current Biology. 14(14): R546-R551.
72. Pfaffl MW (2012). Quantification strategies in real-time polymerase chain reaction. Martin Filion, Hg., *Quantitative real-time PCR in Applied Microbiology*. 53-62.
73. Poetz O, Ostendorp R, Brocks B, Schwenk JM, Stoll D, Joos TO & Templin MF (2005). Protein microarrays for antibody profiling: specificity and affinity determination on a chip. *Proteomics*. 5(9):2402-2411.
74. Poptsova MS, Il'Icheva I A, Nechipurenko DY, Panchenko LA, Khodikov MV, Oparina NY & Grokhovsky SL (2014). Non-random DNA fragmentation in next-generation sequencing. *Scientific reports*. 4.
75. Pu RT & Clark DP (2003). Detection of DNA Methylation. *Acta cytological*. 47(2):247-252.
76. Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R & Turner DJ (2008). A large genome center's improvements to the Illumina sequencing system. *Nature methods*. 5(12):1005-1010.
77. Quail MA, Swerdlow H & Turner DJ (2009). Improved protocols for the illumina genome analyzer sequencing system. *Current protocols in human genetics*. 18-2.
78. Rabbani MA, Maruyama K, Abe H, Khan MA, Katsura K, Ito Y & Yamaguchi-Shinozaki K. (2003). Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. *Plant physiology*, 133(4):1755-1767.
79. Reen DJ (1994). Enzyme-linked immunosorbent assay (ELISA). *Basic Protein and Peptide Protocols*. 461-466.
80. Rosenthal N (1987). Identification of regulatory elements of cloned genes with functional assays. *Methods in enzymology*. 152:704-720.
81. Schefe JH, Lehmann KE, Buschmann IR, Unger T & Funke-Kaiser H (2006). Quantitative real-time RT-PCR data analysis: current concepts and the novel "gene expression's CT difference" formula. *Journal of molecular medicine*. 84(11):901-910.
82. Schena M, Shalon D, Davis RW & Brown PO (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science-New York then Washington*. 467-467.
83. Schnappinger D (2008). Transcriptomics and Transcriptional Regulation. Handbook of Tuberculosis, 213-240.
84. Schuster SC (2007). Next-generation sequencing transforms today's biology. *Nature methods*. 5(1):16.
85. Schuurs AH WM, & Van Weemen BK (1977). Enzyme-immunoassay. *Clinica Chimica Acta*, 81(1):1-40.
86. Shi T, Fillmore TL, Sun X, Zhao R, Schepmoes AA, Hossain M & Moore RJ (2012). Antibody-free, targeted mass-spectrometric approach for quantification of proteins at low picogram per milliliter levels in human plasma/serum. *Proceedings of the National Academy of Sciences*, 109(38):15395-15400.
87. Slade AJ & Knauf VC (2005). TILLING moves beyond functional genomics into crop improvement. *Transgenic research*. 14(2):109-115.
88. Speicher MR, Ballard SG & Ward DC (1996). Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nature genetics*, 12(4):368-375.
89. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H & Timm J (2005). A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, 122(6):957-968.
90. Stephenson J (2008). 1000 Genomes Project. *JAMA*. 299(7): 755-755.
91. Taylor S, Wakem M, Dijkman G, Alsarraj M, & Nguyen M (2010). A practical approach to RT-qPCR—publishing data that conform to the MIQE guidelines. *Methods*. 50(4):S1-S5.
92. Theisen A (2008). Microarray-based comparative genomic hybridization (aCGH). *Nature Education*. 1(1):45.
93. Ungerer MC, Johnson LC & Herman MA (2008). Ecological genomics: understanding gene and genome function in the natural environment. *Heredity*. 100(2):178-183.
94. Valouev A, Johnson DS, Sundquist A, Medina C, Anton E, Batzoglou S & Sidow A (2008). Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nature methods. 5(9):829-834.
95. Van Dijk EL, Auger H, Jaszczyszyn Y & Thermes C (2014). Ten years of next-generation sequencing technology. *Trends in genetics*. 30(9):418-426.
96. Velculescu VE, Zhang L, Vogelstein B & Kinzler KW (1995). Serial analysis of gene expression. *Science*, 270(5235):484.
97. Wall PK, Leebens-Mack J, Chanderbali AS, Barakat A, Wolcott E, Liang H & Ma H (2009). Comparison of next generation sequencing technologies for transcriptome characterization. BMC genomics. 10(1):347.
98. Wang Z, Gerstein M & Snyder M (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics*. 10(1), 57-63.
99. Watson JD (1990). The human genome project: past, present, and future. *Science*, 248(4951):44-50.
100. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A & Gomes X (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 452(7189):872-876.

101. Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU & Wagner L (2003). Database resources of the National Center for Biotechnology. *Nucleic acids research*. 31(1):28-33.

102. White KP (2001). Functional genomics and the study of development, variation and evolution. *Nature Reviews Genetics*, 2(7):528-537.

103. Wilhelm BT & Landry JR (2009). RNA-Seq—quantitative measurement of expression through massively parallel RNA-sequencing. *Methods*, 48(3):249-257.

104. Wold B & Myers RM (2007). Sequence census methods for functional genomics. *Nature methods*, 5(1):19.

105. Wong ML & Medrano JF (2005). Real-time PCR for mRNA quantitation. *Biotechniques*, 39(1):75-88.

106. Wu X, Al Hasan M & Chen JY (2014). Pathway and network analysis in proteomics. *Journal of theoretical biology*. 362:44-52.

107. Yang M, Wu Z & Fields S (1995). Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nucleic Acids Research*, 23(7):1152-1156.

108. Zhang J, Chiodini R, Badr A & Zhang G (2011). The impact of next-generation sequencing on genomics. *Journal of genetics and genomics* 38(3):95-109.

109. Zhao B & Poh CL (2008). Insights into environmental bioremediation by microorganisms through functional genomics and proteomics. *Proteomics*. 8(4):874-881.

How to cite article: Nida Tabassum Khan , Namra Jameel, Maham Jamil Khan. Functional Genomics–Linking Genotype with Phenotype on Genome-wide Scale. Int. J. Appl. Pharm. Sci. Res. 4(1):4-12. doi: https://doi.org/10.21477 ijapsr.4.1.2

Source of Support; Nil **Conflict of Interest:** None declared