IJBPAS, February, 2024, 13(2): 674-687

ISSN: 2277–4998

International Journal of Biology, Pharmacy and Allied Sciences (IJBPAS)

'A Bridge Between Laboratory and Reader'

www.ijbpas.com

ROLE OF BIOINFORMATICS IN MENTAL HEALTH

GARG S AND CHATURVEDI S*

Department of Pharmacology, SGT University Gurugram, Haryana, India 122505

***Corresponding Author: Dr. Sushma Chaturvedi: E Mail: sushma_fphs@sgtuniversity.org**

Received 4th Jan. 2023; Revised 25th March 2023; Accepted 6th June 2023; Available online 1st Feb. 2024

https://doi.org/10.31032/IJBPAS/2024/13.2.7763

ABSTRACT

The bioinformatics presents both distinctive opportunities and difficulties for behavioural neuroscience research. Describing, defining, and differentiating between abstract behavioural processes has proven to be a significant issue, and this is partly due to the biological underpinnings underlying distinct but not completely discrete behavioural entities. Data from many biological systems, data kinds, and scales must be integrated to fully grasp the intricacy of neurobiology and behaviour. It is believed that "true" classifications can be discovered in the shared biology of related disorders and contrasted with the unique biology of separate disorders. This review focusses on the connections between these systems with psychiatric disorders and behavioral traits and the role of bioinformatics in comprehending psychiatric diseases.

Keywords: Neuroscience, Bioinformatics, Behaviour, Psychiatric illness

INTRODUCTION

Neuropsychological, neurobehavioral, and the reduced capacity to comprehend new or complex information are widely viewed as hallmarks of mental illnesses. According to the World Health Organization's 2010 report, major depression will be the mental condition that will have the greatest impact on people's lives by 2030, affecting an estimated 450 million people globally **[1, 2]**. Whatever the original aetiology, it must now be seen as a sign of a behavioural, mental, or biological condition in the individual. Both deviant behaviour and conflicts that are primarily between an individual and society are not mental illnesses, even when they are signs of a person's dysfunction as described above **[3]**. But in addition to being polygenic by nature, psychiatric diseases are also

heavily influenced by the environment. The latter suggests that genomics alone won't be sufficient and that multi-omics data must be gathered, including data from transcriptomics, proteomics, and metabolomics, among others. These data capture various but complementary levels of information, which, when combined, can help identify pathways with deeper biological significance **[4, 5, 6]**.

What is Bioinformatics?

Bioinformatics is the study of biological data collection and understanding using computational and analytical methods. Biology, mathematics, computer science, and physics are all used in this multidisciplinary discipline. Bioinformatics is vital for the management of data in current biology and medicine **[7]**. It is the conceptualization of biology in terms of macromolecules, followed by the application of "informatics" techniques to understand & organise the large-scale knowledge about these compounds. The word "bioinformatics" was created to characterise the activity of research teams that collaborated to address novel problems in the field of biological research **[8, 9]**.

Aim of Bioinformatics

The three main objectives of bioinformatics are listed below. First, arranges data so that researchers can access existing information and submit new things as they are developed, such as the Protein Data Bank for 3D macromolecular structures **[9, 10, 11]**. The creation of materials and tools to assist in data analysis is the second goal. A sequence comparison with other, already defined proteins, for example, might be intriguing. The ultimate goal is to employ these technologies to examine the data and to give physiologically relevant interpretations of the findings **[9, 12].**

Psychiatric disorder classification and challenges

With accordance with DSM 4 (Diagnostic and Statistical Manual of Mental Disorders), An individual's clinically significant behavioural or psychological condition or pattern associated with current discomfort or disability or with a noticeably elevated probability of experiencing death, pain, disability, or a severe loss of freedom **[3]**. According to the DSM 5, it is a condition that is characterized by a clinically significant disturbance in a person's cognition, behavior, or emotion regulation as a result of a breakdown in the biological, psychological, or developmental processes that underpin mental functioning **[13]**. Frameworks for both conceptual and diagnostic thinking have been gradually improved with each DSM iteration **[14, 15, 16, 17]**. According to significant research, it may be more useful to think of emotion, cognition, and behavior as dynamic continuums rather than as discrete categories and constructions of psychiatric

diseases **[14, 18, 19]**. Variable levels of "symptomology" are also produced during development by interactions between a wide range of environmental and biological elements. The best way to incorporate research on the roles of pathophysiological, genetic, and environmental elements into a clinically relevant framework is yet unknown, despite the fact that technology is developing alongside our understanding of complex psychiatric illnesses **[14, 18, 20, 21, 22]**.

Due to the complexity and variability of these conditions, classification of mental and behavioural disorders has proven difficult. As a result, it becomes difficult to name and distinguish distinct entities of behavioural function, which poses problems for diagnostics, research, and therapies **[14]**. Research, diagnosis, and treatment of mental disease are anticipated to be aided by shifting the focus of psychiatric classification from mostly social and subjective externalities to the biological foundation of disorders **[23]**. Behavioral change is achieved by psychotherapy, which is referred to as an "epigenetic medication" in a recent article **[24]**. Researchers are also learning more about the biological impacts of both conventional and newly designed cognitive behavioral therapies **[25, 26, 27, 28]**.

ROLE OF BIOINFORMATICS IN MENTAL HEALTH

By deeply integrating biological information with psychiatric disorders and behavioral traits, bioinformatics tools have the potential to fundamentally alter how we currently comprehend psychiatric diseases. The National Institute of Mental Health is now conducting scientific and clinical research to help identify these traits, and one such attempt is the Research Domain Criteria **[29]**. Our understanding of the biology of complex diseases will be aided by the availability, diversity, and standardization of large-scale data resources together with high-throughput biological assays in basic and clinical research. Although it is a herculean undertaking, compiling neuroscience information is just the first stage in a lengthy process **[30, 31]**. Because psychiatric disorders are so complicated, both basic scientists and practitioners of clinical medicine employ a huge variety of biological research instruments. Modern bioinformatics has advanced far beyond its early roots in sequence analysis to take on the challenge of deep data integration from functional genomics to a variety of other fields, including network modelling, functional and predictive biology, and enabling representation and integration of biological knowledge. One day, thorough and in-depth systems biology studies of the mechanisms underlying mind, emotion, and behavior may be possible thanks to comprehensive

techniques integrating content-rich biological quantification and crucially, well-founded and informative behavioral phenotypes **[32, 33]**.

Methods for Mapping of Biological Substrate to behavioral Function

For the worldwide mapping of biological substrates onto psychological functions and features, genomics and bioinformatics offer new technology and experimental approaches.

1. **Genome annotation**

   Annotating genes involves synthesizing the intron-exon structure from a variety of sources, such as gene prediction, expression data, protein homology, and repetitive elements **[34, 35]**.

There are two steps to it:

- Generation of Evidence
- Synthesis

Evidence is any information that can be used to detect or understand a gene's exon/intron structure. Genome annotation frequently uses a combination of information from closely similar species' whole genome alignments, gene predictions, transcript and protein alignments and repeat masking. The ideal approach to low complexity repeats is soft masking. To do soft masking, lowercase letters are substituted for uppercase letters in the FASTA sequence, while for hard masking, Ns are used instead of letters **[36]**.

Table 1: Evidence Aligners and Assemblers, Gene Predictors

Category	Software	References
Evidence Aligners and Assemblers	BLAST	[36]
	Tophat2	[37]
	GSNAP	[38]
	STAR	[39]
Gene predictors	Genemark	[40]
	SNAP	[35]
	Fgenesh	[41]
	mGene	[42]

Due to the fact that many recently sequenced genomes are not from closely similar species having genomes that have been annotated, they are of significant evolutionary importance. However, several closely similar species are also being sequenced. Several of these are bacterial pathogens with small genomes that are only expected to vary modestly. Whole genome alignment techniques are time- and money-efficient, highly specific methods for these genomes. Examples of programmes that use whole genome multiple alignments to annotate orthologous genes include Mugsy-annotator **[43]** and CONTRAST **[44]**.

Putting this information together to create gene annotations is the annotator's challenging task after evidence has been generated. Annotators would identify the genes based on aligned data and

computational predictions for the genes. In order to create gene models that most nearly matched the exon/intron structure that the data had indicated, the annotators would first identify the genes **[34]**. The small genomic communities of today just lack the significant time and resources needed for manual annotation operations. There are several automated annotation pipelines, and they all accomplish the following: Gene areas are first identified using evidence grouping, and then gene prediction accuracy is increased by using aligned evidence of RNA and Protein. The exon combination that best represents the evidence is subsequently selected by certain pipelines using a combiner or chooser algorithm.

Table 2: Choosers and Combiners, Genome Annotation Pipelines

Category	Software	References
Choosers and Combiners	EVidenceModeler	[45]
	GLEAN	[46]
	JIGSAW	[47]
	Evigan	[48]
Genome Annotation pipelines	MAKER	[49]
	NCBI	[50]
	PASA	[45]
	Ensembl	[51]

2. Functional annotation in mutation screens

Systematic perturbation and functional characterization are a different method to link genes to behavior. In many individual labs, gene manipulation techniques including as mutation, knockout, knockdown, and others have been used extensively. Findings from each of this research are painstakingly filtered to functionally annotate the genome **[52, 53]**. Using mutagenesis **[54]** and targeted deletion **[55]**, several large-scale studies in mice have disrupted vast catalogues of genes. However, it has since been expanded to include mutagenesis of drosophila and various manipulations of the zebrafish. This type of work was initially pioneered in yeast **[56, 57, 58]**. To quickly annotate genes with behaviors, people with these alterations are then routinely examined for behavioral abnormalities. Rapid, large-scale annotation is undertaken, and "phenotypic alleles" are swiftly added to model organism databases using phenotypic screenings that are linked to specific concepts in behavioral ontologies. To uncover relationships among phenotypes, recent research has shown the value of integrating concentrated collections of well-characterized mutants. Yet, collections of models are found for further functional analysis, and this leads to a

variety of different avenues to follow **[59]**.

3. Gene Expression Analysis

While hundreds of genes' expression levels are assessed simultaneously, microarray technology has become a popular technique for genome-wide gene expression profiling **[60]**. The use of microarray technology allows researchers to identify a small number of genes who may be "key players" in the biological phenomena that can be seen. Additionally, it enables researchers to understand changes at the level of molecular networks and pathways as well as the "big picture," find key multi-gene interactions, and observe these changes in context. Because of gene expression profiling, the approach taken in clinical research to find biomarkers is evolving. Nonetheless, there is still work to be done in integrating this technique into genetic medicine. Currently, whole transcriptome profiling is a popular use of next-generation sequencing, but before the method can be extensively used, there are still a number of technical challenges in the processing and analysis of sequencing data **[14]**. Sequence and gene expression array analysis is used to estimate the abundance of a certain sequence in genomic DNA, allowing examination of the number of copies of a single transcript present in various individuals. This method has been used to identify copy number variations in mice and people who exhibit aggression, drunkenness, and anxiety **[61, 62, 63]**. These approaches use "guilt-by-association" and cognate methodologies to locate genes that may share a function with already well-characterized genes known to be involved in disease, with the goal of identifying the cohesive set of genes underpinning shared processes in a strictly empirical manner.

4. Finding the source of genetic variation in behaviour

The mechanisms and makeup of specific comorbidities may be better understood by identifying genes that play a similar function in a number of connected yet seemingly unrelated processes. Overlapping loci can frequently be used to map related features, but since they are so big, it is difficult to refine them to find exact causal polymorphisms **[64, 65]**. A small number of potential regulators are often identified through methods that link behaviour to a gene or region of the genome. Aggregation of functional data makes it easier to find them and rank them. While these genetic techniques help to uncover the biological underpinnings of spontaneous behavioural diversity, they are typically less helpful for understanding the relationships between various illnesses

[66]. The issue is made worse by the fact that the identification of a gene that predicts behavioural variation frequently prompts further, underpowered investigations into the function of the same gene or locus in a variety of other behaviours, leading to a confusing web of associations rather than a comprehensive understanding of the shared function of endophenotypic processes in various behaviours.

5. Trait Correlation, Gene Expression Correlation and System Genetics

Systems genetics is a newly emerging field that combines population genetics of behaviour with systems biology. Systems genetics is based on the genetic correlation of transcripts from one transcript to another, from behaviour to transcript, and among behavioural processes. Biology is genetically altered by naturally occurring genetic polymorphisms. A set of potential network nodes can be found while also identifying causal nodes, which are represented at quantitative trait loci, by building networks solely out of trait correlations. Analysis of expression quantitative trait loci is a novel method that concurrently genotypes individuals using a panel of genetic markers and phenotypes them using DNA microarrays. Because of the spacing between the markers and linkage disequilibrium, it can be difficult to pinpoint the causal factors generating the observed changes in the downstream expression. Each marker could be close to several genes **[67, 68]**. The genomic architecture of gene regulation can be clarified by the use of eQTL mapping. Due to the high degree of dimension in both the data on gene expression and the genomic markers, it is challenging to identify genomic variation in order to understand the variation in expression features. Microarrays and a previously generated linkage map were used to clarify the chromosomal distribution and sex specificity of brain expression QTL in the recently created dwarf (limnetic) and normal (benthic) whitefish **[69]**. Along with being used to restructure regulatory networks, systems genetics was primarily used to refine the identified QTL, discover candidate genes and SNPs, understand gene-environment and gene-gene interactions, find potential regulator genes, separate different QTL, find pleiotropic QTL, and refine the QTL found **[70]**.

CHALLENGES

Recombinant inbred strain utilization is not "fast and easy," but rather takes a considerable lot of planning, infrastructure, and understanding of genetics. The main obstacle was the

ability of mapping resolution because there weren't enough recombinant inbred strains accessible for eQTL testing **[71, 72]**. Although population divergence ultimately results in speciation events, eQTL mapping hasn't been applied to look at the regulatory structure of the genes connected to this process. It can be challenging to identify the causal causes causing the as a result of linkage disequilibrium and marker spacing, variations in the downstream expression have been noted. Many genes may reside nearby each marker.

According to the eQTL measurements alone, every gene at a locus is similar, identifying the true causal gene necessitates additional information. Absence of mechanistic explanation: An relationship between a gene and phenotype often provides little information about the underlying molecular basis of the association **[73,74]**.

Current initiatives show promise for accurately characterising and connecting behavioural traits to different anatomical granularity levels **[75]**. Briefly said, the technology we use to monitor collective biological activity are fast increasing, as are the processing techniques needed to follow numerous measurements over time and place. The information from these integrative investigations is being represented, stored, and shared using bioinformatics tools, enabling a more thorough integration of the central nervous system's activity and how certain parts of that function connect to psychiatric disorders.

CONCLUSION

Technologies for swiftly detecting and defining the function of biological systems in behavioural processes have been made possible by breakthroughs in bioinformatics and complementary high-throughput assessments of brain and behaviour. As a result, fresh molecular targets for research, diagnosis, and treatment have been found. Although a lot of this research is still in its early stages, exciting developments are being made, and practical translation is already taking place. In order to quickly identify relationships like the pleiotropy of gene action, high-throughput biology creates data that can be quickly annotated to these ontologies. The inputs and validating data are derived from actual experimental data, which is frequently gathered utilising high-throughput measurement devices. It is believed that "true" classifications can be discovered in the shared biology of related disorders and contrasted with the unique biology of separate disorders.

Understanding mental illnesses and its underlying processes and characteristics can be defined, categorised, and organised using bioinformatics. Additionally, and this is crucial to any scientific activity, it offers a framework and technology that can be used to test these structures, allowing the categorization system to be falsifiable.

REFERENCES

[1] Zhao W, Yang W, Zheng S, Hu Q, Qiu P, Huang X, Hong X, Lan F. A new bioinformatic insight into the associated proteins in psychiatric disorders. SpringerPlus. 2016 Dec;5:1-8.

[2] Sullivan SG, Hussain R, Threlfall T, Bittles AH. The incidence of cancer in people with intellectual disabilities. Cancer Causes & Control. 2004 Dec;15:1021-5.

[3] Stein DJ, Phillips KA, Bolton D, Fulford KW, Sadler JZ, Kendler KS. What is a mental/psychiatric disorder? From DSM-IV to DSM-V. Psychological medicine. 2010 Nov;40(11):1759-65.

[4] Fernandes BS, Quevedo J, Zhao Z. Fostering precision psychiatry through bioinformatics. Brazilian Journal of Psychiatry. 2022 Mar 1;44:119-20.

[5] Ressler KJ, Williams LM. Big data in psychiatry: multiomics, neuroimaging, computational modeling, and digital phenotyping. Neuropsychopharmacology. 2021 Jan;46(1):1-2.

[6] Paczkowska M, Barenboim J, Sintupisut N, Fox NS, Zhu H, Abd-Rabbo D, Mee MW, Boutros PC, Reimand J. Integrative pathway enrichment analysis of multivariate omics data. Nature communications. 2020 Feb 5;11(1):735.

[7] Bayat A. Science, medicine, and the future: Bioinformatics. BMJ: British Medical Journal. 2002 Apr 4;324(7344):1018.

[8] Fenstermacher D. Introduction to bioinformatics. Journal of the American Society for Information Science and Technology. 2005 Mar;56(5):440-6.

[9] Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics? A proposed definition and overview of the field. Methods of information in medicine. 2001;40(04):346-58.

[10] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic acids research. 2000 Jan 1;28(1):235-42.

[11] Bernstein FC, Koetzle TF, Williams GJ, Meyer Jr EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. Journal of molecular biology. 1977 May 25;112(3):535-42.

[12] Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. Proceedings of the

National Academy of Sciences. 1988 Apr;85(8):2444-8.

[13] Stein DJ, Palk AC, Kendler KS. What is a mental disorder? An exemplar-focused approach. Psychological medicine. 2021 Apr;51(6):894-901.

[14] Chesler EJ, Logan RW. Opportunities for bioinformatics in the classification of behavior and psychiatric disorders. InInternational review of neurobiology 2012 Jan 1 (Vol. 104, pp. 183-211). Academic Press.

[15] Hilsenroth MJ, Ackerman SJ, Blagys MD, Baumann BD, Baity MR, Smith SR, Price JL, Smith CL, Heindselman TL, Mount MK, Holdwick Jr DJ. Reliability and validity of DSM-IV axis V. American Journal of Psychiatry. 2000 Nov 1;157(11):1858-63.

[16] Kotov R, Ruggero CJ, Krueger RF, Watson D, Yuan Q, Zimmerman M. New dimensions in the quantitative classification of mental illness. Archives of general psychiatry. 2011 Oct 3;68(10):1003-11.

[17] Kraemer HC, Kupfer DJ, Clarke DE, Narrow WE, Regier DA. DSM-5: how reliable is reliable enough?. American Journal of Psychiatry. 2012 Jan;169(1):13-5.

[18] Hyman SE. Can neuroscience be integrated into the DSM-V?. Nature Reviews Neuroscience. 2007 Sep;8(9):725-32.

[19] Varga S. Defining mental disorder. Exploring the'natural function'approach. Philosophy, Ethics, and Humanities in Medicine. 2011 Nov;6(1):1-0.

[20] Enoch MA, White KV, Waheed J, Goldman D. Neurophysiological and genetic distinctions between pure and comorbid anxiety disorders. Depression and Anxiety. 2008 May 1;25(5):383-92.

[21] Lawford BR, Young R, Noble EP, Kann B, Ritchie T. The D2 dopamine receptor (DRD2) gene is associated with co-morbid depression, anxiety and social dysfunction in untreated veterans with post-traumatic stress disorder. European psychiatry. 2006 Apr 1;21(3):180-5.

[22] Molina E, Cervilla J, Rivera M, Torres F, Bellón JÁ, Moreno B, King M, Nazareth I, Gutierrez B. Polymorphic variation at the serotonin 1-A receptor gene is associated with comorbid depression and generalized anxiety. Psychiatric genetics. 2011 Aug 1;21(4):195-201.

[23] Craddock N, Owen MJ. The Kraepelinian dichotomy–going, going… but still not gone. The British Journal of Psychiatry. 2010 Feb;196(2):92-5.

[24] Stahl SM. Psychotherapy as an epigenetic 'drug': psychiatric therapeutics target symptoms linked to malfunctioning brain circuits with psychotherapy as well as with drugs. Journal of clinical pharmacy and therapeutics. 2012 Jun;37(3):249-53.

[25] Bryant RA, Felmingham K, Kemp A, Das P, Hughes G, Peduto A, Williams L. Amygdala and ventral anterior cingulate activation predicts treatment response to cognitive behaviour therapy for post-traumatic stress disorder. Psychological medicine. 2008 Apr;38(4):555-61.

[26] Davidson RJ, McEwen BS. Social influences on neuroplasticity: stress and interventions to promote well-being. Nature neuroscience. 2012 May;15(5):689-95.

[27] Huyser C, van den Heuvel OA, Wolters LH, de Haan E, Boer F, Veltman DJ. Increased orbital frontal gray matter volume after cognitive behavioural therapy in paediatric obsessive compulsive disorder. The World Journal of Biological Psychiatry. 2013 May 1;14(4):319-31.

[28] Kobayashi K, Shimizu E, Hashimoto K, Mitsumori M, Koike K, Okamura N, Koizumi H, Ohgake S, Matsuzawa D, Zhang L, Nakazato M. Serum brain-derived neurotrophic factor (BDNF) levels in patients with panic disorder: as a biological predictor of response to group cognitive behavioral therapy. Progress in Neuro-Psychopharmacology and Biological Psychiatry. 2005 Jun 1;29(5):658-63.

[29] Morris SE, Cuthbert BN. Research Domain Criteria: cognitive systems, neural circuits, and dimensions of behavior. Dialogues in clinical neuroscience. 2022 Apr 1.

[30] Crasto CJ, Koslow SH, Marenco L, Nadkarni P, Martone M, Gupta A. Interoperability across neuroscience databases. Neuroinformatics. 2007:23-36.

[31] Martone ME, Gupta A, Ellisman MH. E-neuroscience: challenges and triumphs in integrating distributed data from molecules to brains. Nature neuroscience. 2004 May 1;7(5):467-72.

[32] Akil H, Martone ME, Van Essen DC. Challenges and opportunities in mining neuroscience data. science. 2011 Feb 11;331(6018):708-12.

[33] Markram H. Industrializing neuroscience. Nature. 2007 Jan 11;445(7124):160-1.

[34] Campbell MS, Yandell M. An introduction to genome annotation. Current protocols in bioinformatics. 2015 Dec;52(1):4-1.

[35] Korf I. Gene finding in novel genomes. BMC bioinformatics. 2004 Dec;5(1):1-9.

[36] Korf I, Yandell M, Bedell J. Blast. " O'Reilly Media, Inc."; 2003 Jul 29.

[37] Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology. 2013 Apr;14(4):1-3.

[38] Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics. 2010 Apr 1;26(7):873-81.

[39] Dobin A, Gingeras TR. Mapping RNA-seq reads with STAR. Current protocols in bioinformatics. 2015 Sep;51(1):11-4.

[40] Lomsadze A, Burns PD, Borodovsky M. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. Nucleic acids research. 2014 Sep 2;42(15):e119-.

[41] Solovyev V, Kosarev P, Seledsov I, Vorobyev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome biology. 2006 Aug;7(1):1-2.

[42] Schweikert G, Zien A, Zeller G, Behr J, Dieterich C, Ong CS, Philips P, De Bona F, Hartmann L, Bohlen A, Krüger N. mGene: accurate SVM-based gene finding with an application to nematode genomes. Genome research. 2009 Nov 1;19(11):2133-43.

[43] Angiuoli SV, Dunning Hotopp JC, Salzberg SL, Tettelin H. Improving pan-genome annotation using whole genome multiple alignment. BMC bioinformatics. 2011 Dec;12(1):1-1.

[44] Gross SS, Do CB, Sirota M, Batzoglou S. CONTRAST: a discriminative, phylogeny-free approach to multiple informant de novogene prediction. Genome biology. 2007 Dec;8(12):1-6.

[45] Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome biology. 2008 Sep;9:1-22.

[46] http://aws.amazon.com/ec2/

[47] Allen JE, Salzberg SL. JIGSAW: integration of multiple sources of evidence for gene prediction. Bioinformatics. 2005 Sep 15;21(18):3596-603.k

[48] Liu Q, Mackey AJ, Roos DS, Pereira FC. Evigan: a hidden variable model for integrating gene evidence for eukaryotic gene prediction. Bioinformatics. 2008 Mar 1;24(5):597-605.

[49] Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, Holt C, Alvarado AS, Yandell M. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome research. 2008 Jan 1;18(1):188-96.

[50] Gibney G, Baxevanis AD. Searching NCBI databases using Entrez. Current Protocols in Bioinformatics. 2011 Jun;34(1):1-3.

[51] Curwen V, Eyras E, Andrews TD, Clarke L, Mongin E, Searle SM, Clamp M. The Ensembl automatic gene annotation system. Genome research. 2004 May 1;14(5):942-50.

[52] Knowlton MN, Li T, Ren Y, Bill BR, Ellis LB, Ekker SC. A PATO-compliant zebrafish screening database (MODB): management of morpholino knockdown screen

information. BMC bioinformatics. 2008 Dec;9:1-0.

[53] Smith CL, Goldsmith CA, Eppig JT. The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information. Genome biology. 2005 Jan;6:1-9.

[54] Bult C, Kibbe WA, Snoddy J, Vitaterna M, Swanson D, Pretel S, Li Y, Hohman MM, Rinchik E, Takahashi JS, Frankel WN. A genome end-game: understanding gene function in the nervous system. Nature neuroscience. 2004 May 1;7(5):484-5.

[55] Austin, C. P., Battey, J. F., Bradley, A., Bucan, M., Capecchi, M., Collins, F. S., et al. (2004). The knockout mouse project. Nature Genetics, 36(9), 921–924. http://dx.doi.org/ 10.1038/ng0904-921.

[56] Bedell VM, Westcot SE, Ekker SC. Lessons from morpholino-based screening in zebrafish. Briefings in functional genomics. 2011 Jul 1;10(4):181-8.

[57] Clark KJ, Urban MD, Skuster KJ, Ekker SC. Transgenic zebrafish using transposable elements. InMethods in cell biology 2011 Jan 1 (Vol. 104, pp. 137-149). Academic Press.

[58] Petzold AM, Balciunas D, Sivasubbu S, Clark KJ, Bedell VM, Westcot SE, Myers SR, Moulder GL, Thomas MJ, Ekker SC. Nicotine response genetics in the zebrafish. Proceedings of the National Academy of Sciences. 2009 Nov 3;106(44):18662-7.

[59] Blednov YA, Mayfield RD, Belknap J, Harris RA. Behavioral actions of alcohol: phenotypic relations from multivariate analysis of mutant mouse data. Genes, Brain and Behavior. 2012 Jun;11(4):424-35.

[60] Kim K, Zakharkin SO, Allison DB. Expectations, validity, and reality in gene expression profiling. Journal of clinical epidemiology. 2010 Sep 1;63(9):950-9.

[61] Velez L, Sokoloff G, Miczek KA, Palmer AA, Dulawa SC. Differences in aggressive behavior and DNA copy number variants between BALB/cJ and BALB/cByJ substrains. Behavior genetics. 2010 Mar;40:201-10.

[62] Vu TH, Coccaro EF, Eichler EE, Girirajan S. Genomic architecture of aggression: rare copy number variants in intermittent explosive disorder. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2011 Dec;156(7):808-16.

[63] Williams IV R, Lim JE, Harr B, Wing C, Walters R, Distler MG, Teschke M, Wu C, Wiltshire T, Su AI, Sokoloff G. A common and unstable copy number variant is associated with differences in Glo1 expression and anxiety-like behavior. PloS one. 2009 Mar 6;4(3):e4649.

[64] Flint J. Analysis of quantitative trait loci that influence animal behavior. Journal of neurobiology. 2003 Jan;54(1):46-77.

[65] Milner LC, Buck KJ. Identifying quantitative trait loci (QTLs) and genes (QTGs) for alcohol-related phenotypes in mice. International Review of Neurobiology. 2010 Jan 1;91:173-204.

[66] Lee KW, San Woon P, Teo YY, Sim K. Genome wide association studies (GWAS) and copy number variation (CNV) studies of the major psychoses: what have we learnt?. Neuroscience & Biobehavioral Reviews. 2012 Jan 1;36(1):556-71.

[67] Li H, Deng H. Systems genetics, bioinformatics and eQTL mapping. Genetica. 2010 Oct;138:915-24.

[68] Suthram S, Beyer A, Karp RM, Eldar Y, Ideker T. eQED: an efficient method for interpreting eQTL associations using protein networks. Molecular systems biology. 2008;4(1):162.

[69] Whiteley AR, Derome N, Rogers SM, St-Cyr J, Laroche J, Labbe A, Nolte A, Renaut S, Jeukens J, Bernatchez L. The phenomics and expression quantitative trait locus mapping of brain transcriptomes regulating adaptive divergence in lake whitefish species pairs (Coregonus sp.). Genetics. 2008 Sep 1;180(1):147-64.

[70] Keurentjes JJ, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, Peeters AJ, Vreugdenhil D, Koornneef M, Jansen RC. Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. Proceedings of the National Academy of Sciences. 2007 Jan 30;104(5):1708-13.

[71] Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, Beavis WD, Belknap JK, Bennett B, Berrettini W, Bleich A. The Collaborative Cross, a community resource for the genetic analysis of complex traits Nat Genet 36: 1133–1137.

[72] Flint J, Valdar W, Shifman S, Mott R. Strategies for mapping and cloning quantitative trait genes in rodents. Nature Reviews Genetics. 2005 Apr 1;6(4):271-86.

[73] Rockman MV, Kruglyak L. Genetics of global gene expression. Nature Reviews Genetics. 2006 Nov 1;7(11):862-72.

[74] Schadt EE, Lum PY. Thematic review series: systems biology approaches to metabolic and cardiovascular disorders. Reverse engineering gene networks to identify key drivers of complex disease phenotypes. Journal of lipid research. 2006 Dec 1;47(12):2601-13.

[75] Maynard SM, Mungall CJ, Lewis SE, Imam FT, Martone ME. A knowledge based approach to matching human neurodegenerative disease and animal models. Frontiers in neuroinformatics. 2013 May 14;7:7.