Epigenetic Regulations, Motif and Pathway Identification of Gabaergic Neurotransmitter's Chip Sequence

I. A. Shylesh Murthy[1(✉)] and Preenon Bagchi[1,2,3]

[1] Vasishth Academy of Advanced Studies and Research, Sarvasumana Association, Bengaluru, India
shyleshias@gmail.com
[2] Padmashree Institute of Management and Sciences, Bengaluru, India
[3] MGM Institute of Biosciences and Technology, Aurangabad, India

Abstract. Background: The next generation techniques has revolutionalized the biology era. In this work we are using Chromatin immunoprecipitation (ChiP) technique on the ChIP–seq of Gamma-Amino Butyric Acid (GABA). GABA is a brain-made chemical. It also found in some foods. The effect of GABA in brain is anti-seizure and anti-anxiety. GABA blocks the brain signals, the neurotransmissions and hence there is interest in using GABA as a dietary supplement. But it is taken by mouth, it is able to cross the blood-brain barrier. Hence the effects of GABA supplements on the brain is unknown.

Methodology: GABA is implicated in several neurological disorders like alzheimers, rett symdrome, schizophrenia, etc. In this work, the chip sequence of GABA is retrieved from SRA database, removed the PCR duplicates and finally identified the genome enriched region using the MAC2 call peak tool.

Results: Peak table was annotated to identify the genes and the corresponding pathways was identified from the KEGG pathway. This annotated peak table was used to identify the motifs present in the chip-seq of GABA.

Keywords: Chromatin immunoprecipitation (ChiP) · Gamma-Amino Butyric Acid (GABA) · Neurotransmission · PCR duplicates · MACS2 callpeak · ChiP peak · KEGG pathway · motif

1 Introduction

The advent of Next-generation sequencing (NGS) techniques is utilized in divergent technique research to make an important assortment features of chromatin biology by categorizing genomic loci that are bound by transcription features, occupied by nucleosomes loci [1]. In Chip-seq analysis mapping of the DNA-protein interfaces occurs at very high decree and hence it gives proactive quality-control and appropriate dataset analytics which help in take important decisions resulting in meaningful analysis. This High resolution NGS technique resolves the way for genome-wide profiling and

© The Author(s) 2023
R. Somashekhar et al. (Eds.): ICBDS 2022, AHSR 58, pp. 121–131, 2023.
https://doi.org/10.2991/978-94-6463-164-7_10

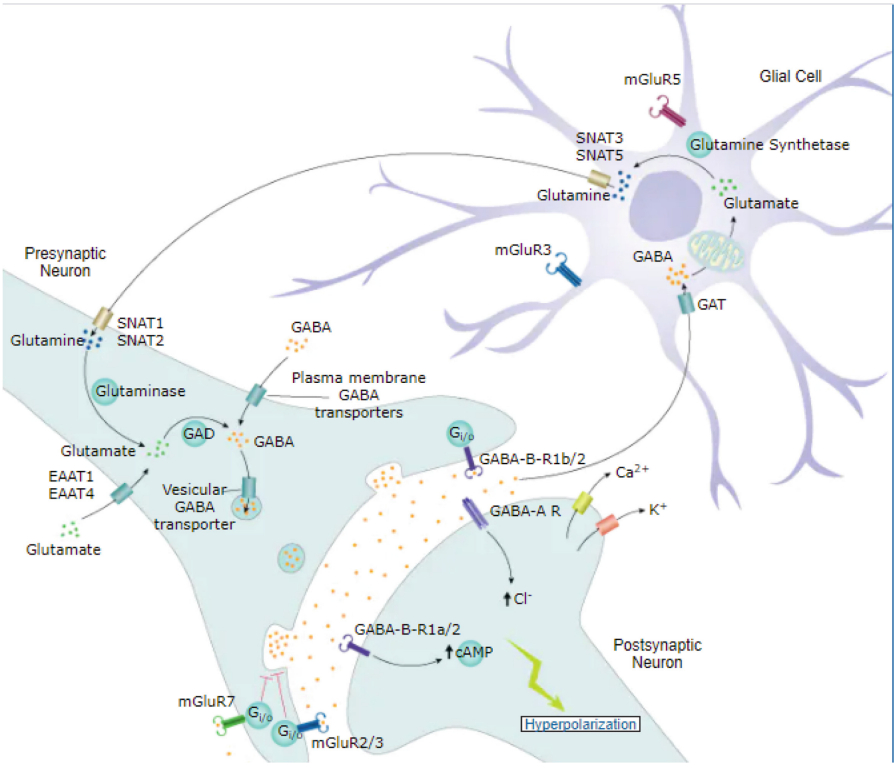


Fig. 1. The GABA pathway [7]

tremendous progress of DNA-binding proteins, histone modifications, or nucleosomes [2–4].

We have taken neurotransmitter Gamma-Amino Butyric Acid (GABA)ergic ChIP sequence for our present study. The long-term stability and function of neuronal networks is dependent on a maintained balance between excitatory and inhibitory synaptic transmission that functions as the primary inhibitory neurotransmitter by inhibiting nerve transmission and reduces neuronal excitability in the central nervous system (CNS) (Fig. 1) [5, 6].

The GABA released or liberated from presynaptic terminals mediate the GABAergic signaling, through a specific actions by the neurotransmission of GABA. This happens by using a class of GABA transporters (GATs) which is followed by the actions of GABA which plays a very important and very a crucial role in controlling both neuronal inhibitory and excitatory states, as well as in wide range of developmental with behavioral patterns in the brain [8–10]. Any fluctuations in GABA and GABAergic signaling are implicated in various neurological disorders like autism spectrum disorder (ASD), Rett's syndrome, Prader-Willi Syndrome (PWS) and schizophrenia (SZ) [11, 12].

2 Materials and Methods

The primary aim of the current research is to provide a detailed overview of epigenetic regulations of GABAergic neurotransmission and throw a light on the pathways involved. Galaxy tutorial by Lauren Mills, Analyzing ChIP-Seq Data in Galaxy is used to analyze *Mus musculus* GABAergic CB1-KO sequence having SRA accession number SRR13960849 and SRR13960850 from SRA database taken for this work. We mapped the reads using Bowtie2 [13, 14]. Using Collect Alignment Summary Metrics tool [15] we take the summary of our alignment. Next using RmDup [16] we remove PCR duplicates and Collect Alignment Summary Metrics tool was re-run. Finally, using MACS2 callpeak [17, 18] we identify peaks from alignment results. Using Peak calling we identify areas in our genome that have been enriched with our aligned reads, these areas are those where protein interacts with DNA [19].

Next, we annotate our peaks table to take top 100 most significant peaks and identify the genes overlapping with these peaks. The pathways of these genes were identified from KEGG pathways. Finally, we identify the motifs using SeqPos motif analysis tool [20]. Biological sequence motifs are short conserved sequence pattern associated with distinct functions that usually represents important structural or functional features [21].

3 Results and Discussion

Mus musculus chip-sequences with SRA accession number SRR13960849 and SRR13960850 were mapped using *Mus musculus* ref seq using Bowtie2. The Bowtie2 mapping output is given in Fig. 2 for SRR13960849 and Fig. 3 for SRR13960850.

We use the tool Collect Alignment Summary Metrics tool take the summary of our mapping done above. Table 1 contains the alignment summary for SRR13960849 and Table 2 for SRR13960850.

Next, we removed the PCR duplicates using RmDup. Table 3 & 4 contains the Alignment summary for SRR13960849 and SRR13960850 and removing PCR duplicates.

As per the alignment summary (Table 3, 4), we see that the reads are less post RmDup which implies that the duplicate reads are removed. Next, we use MAC2 call peak tool to identify areas in the genome that are enriched with the aligned reads. Model-based Analysis of ChIP-seq (MACS) is a commonly used tool for identifying transcription factor binding sites. The algorithm confines the influence of genome complexity to evaluate the significance of enriched ChIP regions. This tool improves the spatial resolution of binding sites by combining the information of both sequencing tag position and orientation [22]. Here, MACS is used along with a control sample (SRR13960849) which increases specificity of the peak calls (Fig. 4). MACS2 models the distance between the paired forward and reverse strand peaks and uses 1000 enriched regions to model the distance between the forward and reverse strand peaks [23].

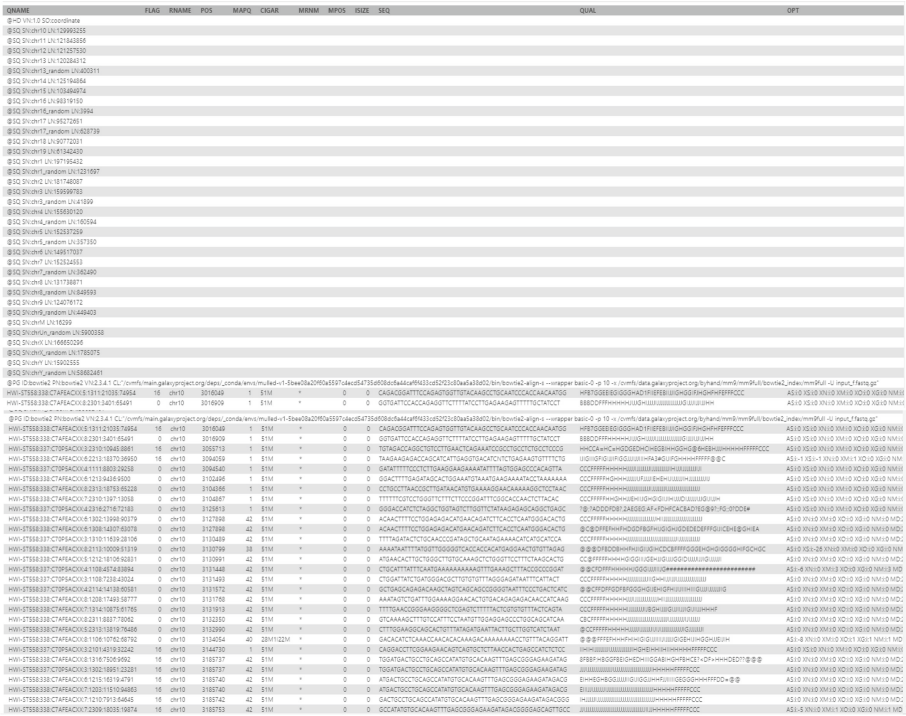


Fig. 2. Mapping output for SRR13960849

The cross-correlation metric diagram given in Fig. 5 is worked out as the Pearson's linear correlation joining the Crick strand and the Watson strand and this metric standard usually produces two peaks when this cross-correlation is plotted against the shift value, one corresponding to the read length ("phantom" peak) and the other average fragment length of the library.

The absolute and relative height of the above said peaks are useful in determining of the success of any ChIP-seq experiment. A high-quality immunoprecipitation is characterized by a ChIP peak that is which should be higher than the "phantom" peak and very small or no ChiP peak is seen in failed experiments. Our results show high and only ChiP peak and no "phantom" peak which clearly says our results contain a high-quality immunoprecipitation [17].

Next, from the annotation results our peaks table by taking top 100 most significant peaks, we identify the genes overlapping with these peaks which are given in Table 5. The pathways of these genes were identified from KEGG pathways (Table 6).

Finally, we identify the motifs present in our GABA ChiP-seq. We used used SeqPos motif analysis tool. The file, top 100 most significant peaks in bed format was selected for motif identification (Table 7, 8).

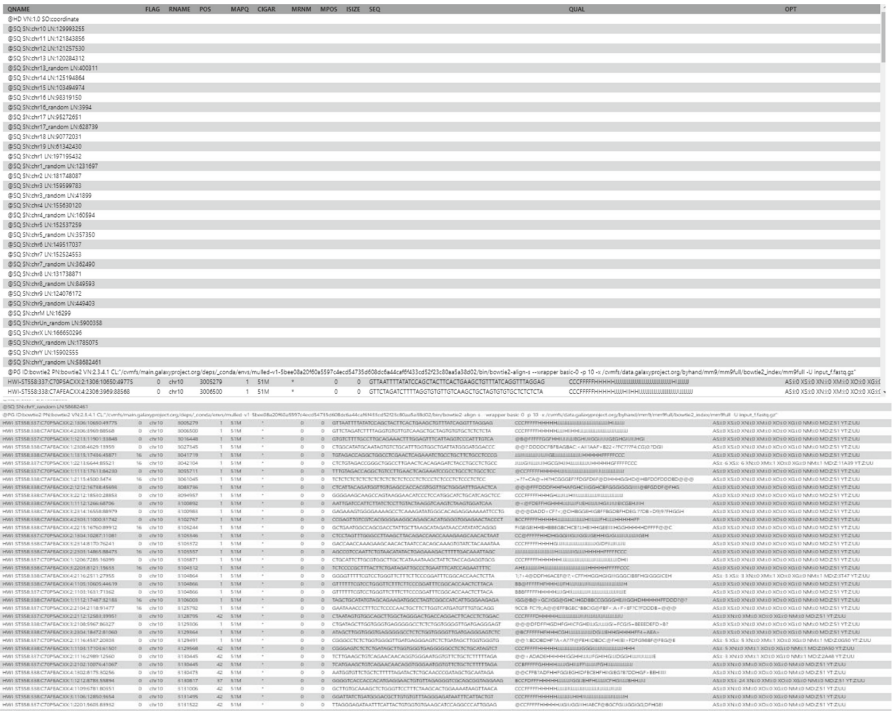


Fig. 3. Mapping output for SRR13960850

Table 1. Alignment summary for SRR13960849

CATEGORY	TOTAL_READS	PF_READS	PF_READS_ALIGNED	PCT_PF_READS_ALIGNED	PF_ALIGNED_BASES	PF_HQ_MEDIAN_MISMATCHES
UNPAIRED	43468872	43468872	40442044	0.930368	2060634150	0.003736

PF_MISMATCH_RATE	PF_HQ_ERROR_RATE	PF_INDEL_RATE	MEAN_READ_LENGTH	READS_ALIGNED_IN_PAIRS	PCT_READS_ALIGNED_IN_PAIRS	PCT_ADAPTER
0.000501	51	0	0	0	0	0.000008

Table 2. Alignment summary for SRR13960850

CATEGORY	TOTAL_READS	PF_READS	PF_READS_ALIGNED	PCT_PF_READS_ALIGNED	PF_ALIGNED_BASES	PF_HQ_MEDIAN_MISMATCHES
UNPAIRED	43309917	43309917	40251830	0.929391	2050849871	0

PF_MISMATCH_RATE	PF_HQ_ERROR_RATE	PF_INDEL_RATE	MEAN_READ_LENGTH	READS_ALIGNED_IN_PAIRS	PCT_READS_ALIGNED_IN_PAIRS	PCT_ADAPTER
0.005489	0.003488	0.000524	51	0	0	0.000009

Table 3. Alignment summary for SRR13960849 post RmDup

CATEGORY	TOTAL_READS	PF_READS	PF_READS_ALIGNED	PCT_PF_READS_ALIGNED	PF_ALIGNED_BASES	PF_HQ_MEDIAN_MISMATCHES
UNPAIRED	17130594	17130594	14103766	0.823309	718686483	0

PF_MISMATCH_RATE	PF_HQ_ERROR_RATE	PCT_PF_READS_IMPROPER_PAIRS	BAD_CYCLES	STRAND_BALANCE	PCT_CHIMERAS	PCT_ADAPTER
0.005145	51	0.003344	0.000468	0.495284	0	0.00002

Table 4. Alignment summary for SRR13960850 post RmDup

CATEGORY	TOTAL_READS	PF_READS	PF_READS_ALIGNED	PCT_PF_READS_ALIGNED	PF_ALIGNED_BASES	PF_HQ_MEDIAN_MISMATCHES
UNPAIRED	17724623	17724623	14666536	0.827467	747351950	0

PF_MISMATCH_RATE	PF_HQ_ERROR_RATE	PF_INDEL_RATE	MEAN_READ_LENGTH	STRAND_BALANCE	PCT_CHIMERAS	PCT_ADAPTER
0.005025	0.003231	0.000477	51	0.49489	0	0.000022

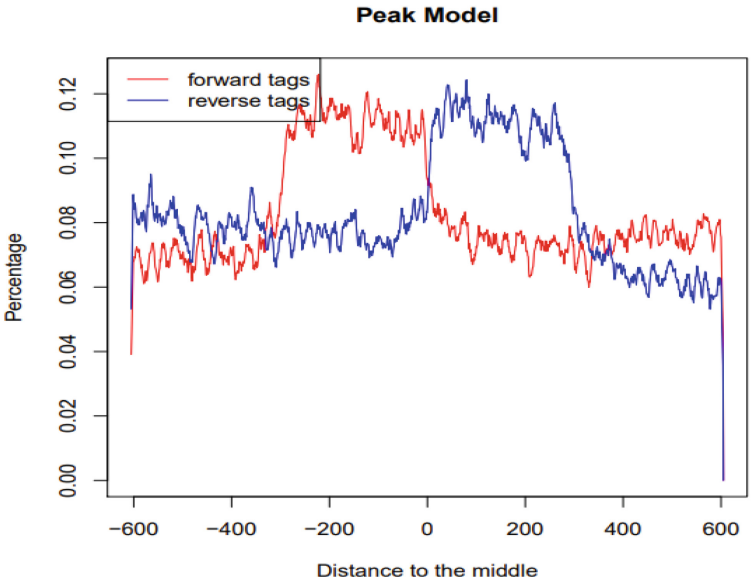


Fig. 4. Peak Model in graphical format

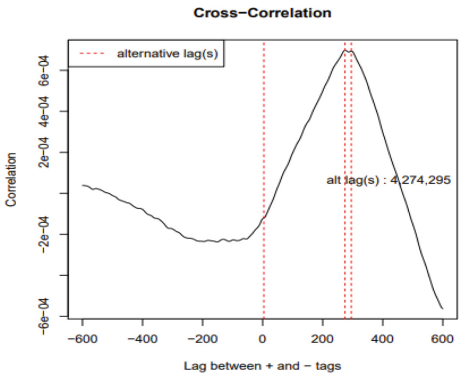
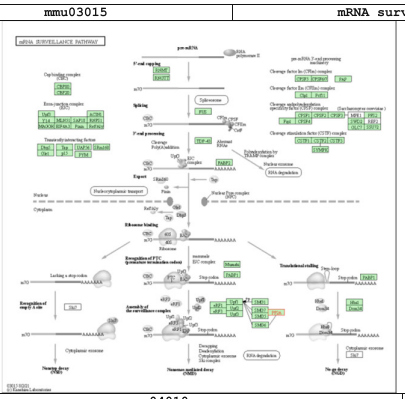
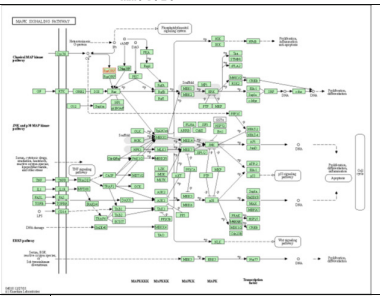
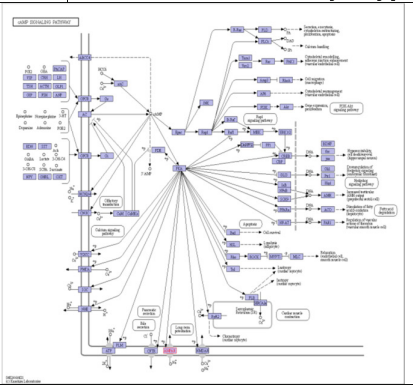


Fig. 5. Cross-Correlation Metric

Table 5. Genes identified

1.	NM_001161362
2.	NM_011245
3.	NM_001357743
4.	NM_133733
5.	NM_001284328
6.	NM_019796
7.	NM_001311113
8.	NM_001113180
9.	NM_019691
10.	NM_033607
11.	NM_008927

Table 6. Pathways for the genes identified

NM_001161362	**mmu03015**	**mRNA surveillance pathway**
		
NM_011245	**mm04010**	**MAPK signaling pathway**
		
NM_001357743	**mm04010**	**MAPK signaling pathway**
NM_133733	09190 Not Included in Pathway or Brite	CXADR-like membrane protein (Clmp), mRNA
NM_001284328	No pathway	synaptotagmin binding, cytoplasmic RNA interacting protein (Syncrip), transcript variant 3, mRNA
NM_019796	No pathway	Mus musculus synaptotagmin binding, cytoplasmic RNA interacting protein (Syncrip), transcript variant 2, mRNA
NM_001311113	No pathway	Mus musculus synaptotagmin binding, cytoplasmic RNA interacting protein (Syncrip), transcript variant 4, mRNA
NM_001113180	No pathway	Mus musculus glutamate receptor, ionotropic, AMPA4 (alpha 4) (Gria4), transcript variant 2, mRNA
NM_019691.5	**ko04024**	**cAMP signaling pathway**
		
NM_033607	No pathway	Mus musculus ubiquitin carboxyl-terminal esterase L4 (Uchl4), mRNA
NM_008927	**ko01521**	**EGFR tyrosine kinase inhibitor resistance**

(*continued*)

Table 6. (*continued*)

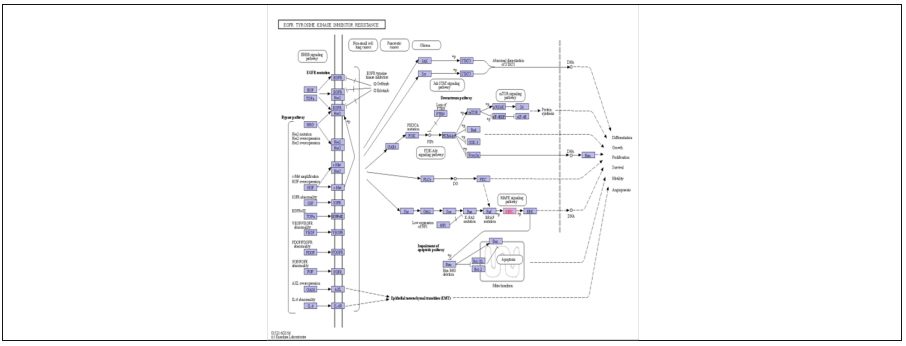
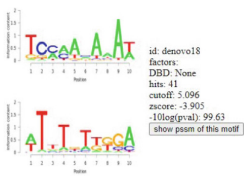
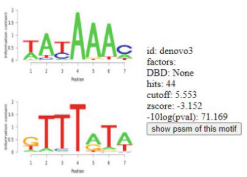


Table 7. Motifs identified [24]

clusters	collapsed_id	factor	DNA binding domain	hits	cutoff	zscore	-10*log(pval)	similarity to top	mean_position
1	denovo18		None	41	5.096	-3.905	99.63		-0.192
2	denovo3		None	44	5.553	-3.152	71.169		-0.159

Table 8. Motifs detected [24]

collapsed_id		hits	cutoff	z-score	-10*log(pval)	mean_position
DENOVO 18	 id: denovo18 factors: DBD: None hits: 41 cutoff: 5.096 zscore: -3.905 -10log(pval): 99.63 show pssm of this motif	41	5.096	- 3.095	99.63	-0.192
DENOVO : 3	 id: denovo3 factors: DBD: None hits: 44 cutoff: 5.553 zscore: -3.152 -10log(pval): 71.169 show pssm of this motif	44	5.553	-3.1552	71.169	-0.159

4 Conclusion

A high-quality immunoprecipitation ChIP peak is detected in our results which imply high concentration of DNA binding proteins. Further, top DNA binding protein genes were identified and corresponding pathways was identified from KEGG pathway. The mRNA surveillance pathway, MAPK signaling pathway, cAMP signaling pathway and EGFR tyrosine kinase inhibitor resistance were detected. Further, motifs were detected from the ChIP peak.

References

1. Meyer CA, Liu XS. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. Nat Rev Genet. 2014;15(11):709-721.
2. Ambrosini, G., Dreos, R., Kumar, S. et al. The ChIP-Seq tools and web server: a resource for analyzing ChIP-seq and other types of genomic data. BMC Genomics 2016, 17, 938.
3. Deliard S, Zhao J, Xia Q, Grant SF. Generation of high quality chromatin immunoprecipitation DNA template for high-throughput sequencing (ChIP-seq). J Vis Exp. 2013 Apr 19;(74):50286.
4. Furey TS. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nat Rev Genet. 2012;13(12):840-852.
5. Tatti R, Haley MS, Swanson OK, Tselha T, Maffei A. Neurophysiology and Regulation of the Balance Between Excitation and Inhibition in Neocortical Circuits. Biol Psychiatry. 2017;81(10):821-831.
6. Gatto CL, Broadie K. Genetic controls balancing excitatory and inhibitory synaptogenesis in neurodevelopmental disorder models. Front Synaptic Neurosci. 2010;2:4. Published 2010 Jun 7.
7. https://www.rndsystems.com/pathways/synaptic-neurotransmission-pathways-gabaergic-inhibition
8. Savtchenko, L., Megalogeni, M., Rusakov, D. et al. Synaptic GABA release prevents GABA transporter type-1 reversal during excessive network activity. Nat Commun 6, 6597 (2015).
9. Wu C, Sun D. GABA receptors in brain development, function, and injury. Metab Brain Dis. 2015;30(2):367-379.
10. Shrestha S, Offer S, M: Epigenetic Regulations of GABAergic Neurotransmission: Relevance for Neurological Disorders and Epigenetic Therapy. Med Epigenet 2016;4:1-19.
11. Kim YS, Yoon BE. Altered GABAergic Signaling in Brain Disease at Various Stages of Life. Exp Neurobiol. 2017;26(3):122-131.
12. Horder J, Petrinovic MM, Mendez MA, Bruns A, Takumi T, Spooren W, Barker GJ, Künnecke B, Murphy DG. Glutamate and GABA in autism spectrum disorder-a translational magnetic resonance spectroscopy study in man and rodent models. Transl Psychiatry. 2018 May 25;8(1):106.
13. Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology, 10(3), R25.
14. Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4), 357–359.
15. Institute, B. (n.d.). Picard. Broad Institute, GitHub repository. GitHub. Retrieved from http://broadinstitute.github.io/picard/
16. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), 2078–2079. Retrieved from https://doi.org/10.1093/bioinformatics/btp352.
17. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. (2008) Model-based Analysis of ChIP-Seq (MACS), Genome Biology, 2008;9(9):R137.
18. Feng, J., Liu, T., Qin, B., Zhang, Y., & Liu, X. S. (2012). Identifying ChIP-seq enrichment using MACS. Nature Protocols, 7(9), 1728–1740.
19. Jiang S, Mortazavi A. Integrating ChIP-seq with other functional genomics data. Brief Funct Genomics. 2018;17(2):104-115. doi:https://doi.org/10.1093/bfgp/ely002.
20. Liu T, Ortiz JA, Taing L, et al. Cistrome: an integrative platform for transcriptional regulation studies. Genome Biol. 2011;12(8):R83. Published 2011 Aug 22. doi:https://doi.org/10.1186/gb-2011-12-8-r83

21. Mohamed SAEH, Elloumi M and Julie D. Thompson (December 14th 2016). Motif Discovery in Protein Sequences, Pattern Recognition - Analysis and Applications, S. Ramakrishnan, IntechOpen, DOI: https://doi.org/10.5772/65441. Available from: https://www.intechopen.com/chapters/52534
22. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, Chen Y, DeSalvo G, Epstein C, Fisher-Aylor KI, Euskirchen G, Gerstein M, Gertz J, Hartemink AJ, Hoffman MM, Iyer VR, Jung YL, Karmakar S, Kellis M, Kharchenko PV, Li Q, Liu T, Liu XS, Ma L, Milosavljevic A, Myers RM, Park PJ, Pazin MJ, Perry MD, Raha D, Reddy TE, Rozowsky J, Shoresh N, Sidow A, Slattery M, Stamatoyannopoulos JA, Tolstorukov MY, White KP, Xi S, Farnham PJ, Lieb JD, Wold BJ, Snyder M. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 2012 Sep;22(9):1813-31. doi: https://doi.org/10.1101/gr.136184.111. PMID: 22955991; PMCID: PMC3431496.
23. Feng J, Liu T, Zhang Y. Using MACS to identify peaks from ChIP-Seq data. Curr Protoc Bioinformatics. 2011;Chapter 2:Unit2.14–2.14. doi:https://doi.org/10.1002/0471250953.bi0214s34
24. Jiang P, Singh M. CCAT: Combinatorial Code Analysis Tool for transcriptional regulation. Nucleic Acids Res. 2014 Mar 1;42(5):2833-47.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.