GMR
The Original

COMPARATIVE GENOMIC CHARACTERIZATION OF CONSERVED REGULATORY ELEMENTS ACROSS SPECIES

Shanmuga Priyan[1], Elansuriyan Sambandam[2], Dr. Umema Ahmed[3], Prathiba S[4], Dr. Prithvi kumar H. Ahir[5], Sahil Suri[6], Prithiviraj Nagarajan[7], Dr. S. Jagadeeswari[8]

[1]Professor, Department of Pharmacology, Meenakshi Medical College Hospital & Research Institute, Meenakshi Academy of Higher Education and Research
[2]Assistant Professor, School of Physiotherapy, Sri Balaji Vidyapeeth University, India, Email: elansuriyan2893@gmail.com, ORCID: https://orcid.org/0009-0006-9101-2921
[3]Associate Professor, Department of FBAS (Forensic Science), Vivekananda Global University, Jaipur, India, Email: umema.a@vgu.ac.in ORCID: https://orcid.org/0000-0002-3754-2941
[4]Lecturer, Department of Pharmacology, Meenakshi Ammal Dental College and Hospital, Meenakshi Academy of Higher Education and Research
[5]Medical Officer, Faculty of Homoeopathy, Gokul Global University, Sidhpur, Gujarat, India, Email: phahir.bhms@gokuluniversity.ac.in ORCID: https://orcid.org/0009-0007-6772-9188
[6]Centre of Research Impact and Outcome, Chitkara University, Rajpura – 140417, Punjab, India, Email: sahil.suri.orp@chitkara.edu.in ORCID: https://orcid.org/0009-0000-4917-9337
[7]Department of Medical Biotechnology, Aarupadai Veedu Medical College and Hospital, Vinayaka Missions Research Foundation (Deemed to be University), Puducherry, India, Email: prithivinaga@gmail.com, ORCID: https://orcid.org/0000-0002-9428-6243
[8]Professor, Paediatrics, Sree Balaji Medical College and Hospital, Bharath Institute of Higher Education and Research, ORCID: https://orcid.org/0000-0002-8548-1207

ABSTRACT
Conserved regulatory elements (CREs) are important in the regulation of gene expression, genomic integrity in a wide range of species. Promoters, enhancers, silencers, and insulators are just few among these components that are usually maintained by evolution because of their essential functional importance. It has spawned the use of comparative genomics as an effective method of locating and describing such conserved regions, which then allows further understanding of their evolutionary background and biological functions. The review under discussion is a thorough study of the conserved elements of regulation in species, including classification, conservation throughout evolution, and functional value. It also discusses the major computational and experimental strategies involved in identifying them, such as sequence alignment strategies, phylogenetic footprinting and high throughput functional genomics methods, including ChIP-seq and ATAC-seq. Also, the latest developments in the functional validation with the help of genome editing technologies are mentioned. Other usage of conserved regulatory element analysis in evolutionary biology, developmental regulation, and disease genomics are also majorly noted in the review. Although there has been great improvement the following issues still persist: context-dependent functionality, incomplete genomic annotations and cross-species variability. The directions of the future that will focus on integrative multi-omics and artificial intelligence-driven approaches are described to deepen the knowledge of regulatory conservation. On the whole, this review highlights the significance of comparative genomic characterization in the demystification of the process of gene regulation in the context of various species.

KEYWORDS: Comparative genomics, conserved regulatory elements, gene regulation, evolutionary conservation, functional genomics, cross-species analysis

1. INTRODUCTION

The control of genes is a basic biological process that controls the expression of genetic information in a specific spatial and temporal pattern, thus, allowing cells to execute specific functions and react to a great variety of environmental signals (Andersson and Sandelin, 2020; Heinz et al., 2015). Protein-coding genes represent a rather small proportion of the genome, and the non-coding parts of the genome are a large portion of the genomic DNA, which has crucial and highly sophisticated regulatory functions (Shlyueva et al., 2014; Pennacchio et al., 2013). One of them, the so-called conserved regulatory elements (CREs), such as promoters, enhancers, silencers, and insulators, plays a major role in the regulation of gene expression at various levels (Long et al., 2016; Andersson et al., 2014). Those factors are dynamic and interact with transcription factors and chromatin-modifying complexes to control complex transcriptional programs (Heinz et al., 2015; Thurman et al., 2012). Notably, most of these regulation sequences are preserved across the species because of intense evolutionary forces and this points to their functional importance and the necessity of preserving biological wholeness (Lowe et al., 2011; Schmidt et al., 2010). As next-generation sequencing technologies and large-scale genomic projects have developed, it is now becoming increasingly possible to identify and compare conserved non-coding regions across a broad set of organisms, making comparative genomics increasingly important to gain insights into the general regulatory structure of genomes (Kundaje et al., 2015; Zerbino et al., 2015). Evolutionary conservation of regulatory elements, which is studied in various species, brings important information both on the principles of gene regulation and on the evolution of genomes (Villar et al., 2015; Lowe et al., 2011). Their conservation over the evolutionary time is a strong indicator that they likely play a central role in sustaining the important biological processes of development, cell differentiation, and physiological adaptation (Nord et al., 2013; Long et al., 2016). Comparative genomic studies allow the investigation of functionally significant regulatory domains that would not be easily obvious based on the study of one genome (Shlyueva et al., 2014; Andersson et al., 2014). Moreover, these methods enable finding common transcription factor binding sites and regulatory motifs, and thus help to understand the intricate system of genes regulation better (Gusmao et al., 2016; Ernst and Kellis, 2012). The implications of this growing body of knowledge in the evolutionary biology field are widespread because they are useful in explaining divergence of species and why some traits are conserved (Lowe et al., 2011). Moreover, it can be also of great importance in biomedical studies, where modifications or mutation of regulatory factors are becoming linked to an extensive collection of genetic diseases (Thurman et al., 2012; Rada-Iglesias et al., 2012). Further, understanding of preserved regulatory processes finds practical uses in the agricultural and biotechnology field with the selective use of the regulatory factors being utilized to enhance crop yield, stress tolerance, and overall crop productivity. The purpose of this review is to critically summarize conserved regulatory elements in terms of comparative genomic context and especially focusing on their classification, evolutionary conservation and functional representation in different species. It also discusses the key computational and experimental methodologies to discover and describe these elements, such as sequence alignment technologies, phylogenetic footprinting, and high-throughput functional genomics technologies (Arnold et al., 2013; Buenrostro et al., 2015; Kvon, 2015). Besides that, the review also indicates the increased role of computational and data-driven methods in better predicting and annotating conserved regulatory elements, especially in the case of large-scale and multi-species genomic data (Zerbino et al., 2015; Gusmao et al., 2016). Important applications in other important areas like disease genomics, developmental biology and evolutionary studies are also discussed in the review. Regardless of the strides that have been made in this field, a number of challenges still exist, among them being, unfinished annotation of genomes, regulation activity that can be context-dependent and inter-species variability. To deal with these challenges, integrative strategies need to be employed to integrate multi-omics data and using sophisticated computational frameworks. It is advisable that in future directions in this field the more strategic approach will be based on data-driven and computational approaches to advance the prediction and functional interpretation of regulatory factors, which will eventually lead to a better understanding of gene regulation in various biological systems.

2. LITERATURE REVIEW

Analysis of conserved regulatory elements (CREs) has become a popular area of study that has developed alongside the development of high-throughput sequencing technology and comparative genomics (Kundaje et al., 2015; Thurman et al., 2012). Initial pioneering work had determined that non-coding genome regions may be highly evolutionarily conserved indicating that they may be functionally relevant in the regulation of genes. Among the most epigenetic contributions ever offered by Bejerano et al. (2004) was establishment of ultraconserved elements in vertebrates, which proved that some non-coding sequences are nearly the same even in distantly related species. This finding changed the emphasis of genomics to regulatory DNA instead of protein-coding genes with particular interest in conserved non-coding elements (CNEs). Thereafter, there were large-scale efforts like the ENCODE Consortium (2012) to systematically map functional elements across the human genome, which found that a large fraction of conserved sequences are regulatory regions such as transcription factor binding sites and chromatin-accessible regions (Kundaje et al., 2015). Likewise, the works of Pennacchio et al. (2006) and Visel et al. (2007) consisted of experimental confirmation of conserved enhancers and showed their involvement in the process of tissue-specific gene expression and development (Pennacchio et al., 2013; Nord et al., 2013). These observations supported the idea that evolutionary conservation is one of the powerful predictors of regulatory purpose. Table 1 provides a summary of important studies, methodologies and their contributions to the knowledge of conserved regulatory elements. Comparative genomic approaches have been used to be at the forefront of the discovery of CREs in cross-species (Lowe et al., 2011; Schmidt et al., 2010). According to Hardison (2000), phylogenetic footprinting allows revealing conserved regulatory regions through aligning the orthologues of several species. Also, probabilistic sequence conservation models created by Siepel et al. (2005) have spawned tools like phastCons and phyloP that are commonly employed to detect conserved elements on a genome- wide basis. The combination of the functional genomics data, such as ChIP-seq and ATAC-seq, has additionally improved the precision in the recognition of the CREs, as it gives the information about chromatin accessibility and transcription factor binding (Buenrostro et al., 2015; Gusmao et al., 2016). In spite of these developments, there are a number of gaps in research. The majority of research is strongly constrained to model organisms like humans, mice and zebrafish, which restricts the learning on regulatory conservation in non-model organisms. Moreover, not every conserved sequence is functional and vice versa, some regulatory elements are functionally conserved, even though the sequences have diverged. The context-dependence of regulatory activity which differs from cell types, developmental stages and environmental conditions is also a major challenge. Thus, an increase in the demand of integrating methods that unite comparative genomics to multi-omics and enhanced computational frameworks to enhance the detection and functional clarification of conserved regulatory components is rising.

Table 1. Summary of Key Studies on Conserved Regulatory Elements Across Species

Author (Year)	Species Studied	Methodology	Key Findings	Limitations
Bejerano et al. (2004)	Vertebrates (human, mouse, rat)	Comparative genomics, sequence alignment	Identified ultraconserved non-coding elements with high evolutionary conservation	Limited experimental validation of function
ENCODE Consortium (2012)	Human	High-throughput functional genomics (ChIP-seq, DNase-seq)	Genome-wide mapping of functional regulatory elements	Focus primarily on human genome
Pennacchio et al. (2006)	Mammals	Comparative genomics, enhancer assays	Demonstrated functional activity of conserved enhancers in development	Limited tissue and species coverage
Visel et al. (2007)	Mouse, human	VISTA enhancer analysis, transgenic assays	Identified tissue-specific conserved enhancers	Experimental validation limited to selected regions
Hardison (2000)	Vertebrates	Phylogenetic footprinting	Introduced methods for detecting conserved regulatory sequences	Dependent on sequence conservation only
Siepel et al. (2005)	Multiple species	Evolutionary models (phastCons, phyloP)	Developed tools for genome-wide identification of conserved elements	Functional interpretation remains challenging

3. Conserved Regulatory Elements: Classification, Evolutionary Basis and Functional Significance

CREs are important elements of the genome that are conserved and regulate the expression of genes in various biological conditions and species. Such factors are promoters, enhancers, silencers and insulators, and all of them have a unique role in the regulation of transcriptional activity. Promoters are usually positioned near transcription start sites and are needed to activate the expression of the gene whereas enhancers may be positioned at distant sites to amplify the efficiency of transcription. Silencers play a role by suppressing the activity of genes and insulators assist in keeping the genomic structure correctly arranged by excluding unsuitable interactions between the regulators. Their adequacy is highlighted in the conservation of these elements through species, which points towards the need to sustain important biological processes. The processes of evolutionary conservation of regulatory elements are conferred by the pressure to maintain valid functional sequences. Comparative genomic analysis has shown that conserved non-coding regions have a high probability to match to regulatory elements which are highly essential in development and cellular processes. These elements are preserved across species through the action of sequence homology, structural conservation and functional conservation. Although there are elements of regulation that have high sequence similarity, there are those that retain their functionality with a divergence in sequence, which underscores the complexity of regulatory evolution. These conserved regions have been commonly identified by use of phylogenetic footprinting, and multi-species genome alignment that facilitates researchers to differentiate between functional and non-functional sequences of the genome. Conserved regulatory elements like promoters and enhancers are also found to be aligned across species showing their conservation through evolution even when species become diverged, as shown in Figure 1. The functional importance of the conserved regulatory elements is spread across a wide range of biological processes including regulation of gene expression, developmental biology and disease pathology. All these components constitute complex regulatory systems through their interactions with transcription factors and epigenetic modifiers, hence affecting the cellular differentiation and development of an organism. Characteristic of many genetic disorders and diseases, disruptions or mutations in conserved regulatory regions have been linked to these conditions and hence their clinical significance. Moreover, conserved regulatory elements are very important in evolutionary biology since they help in conservation and divergence of phenotypes among the species.

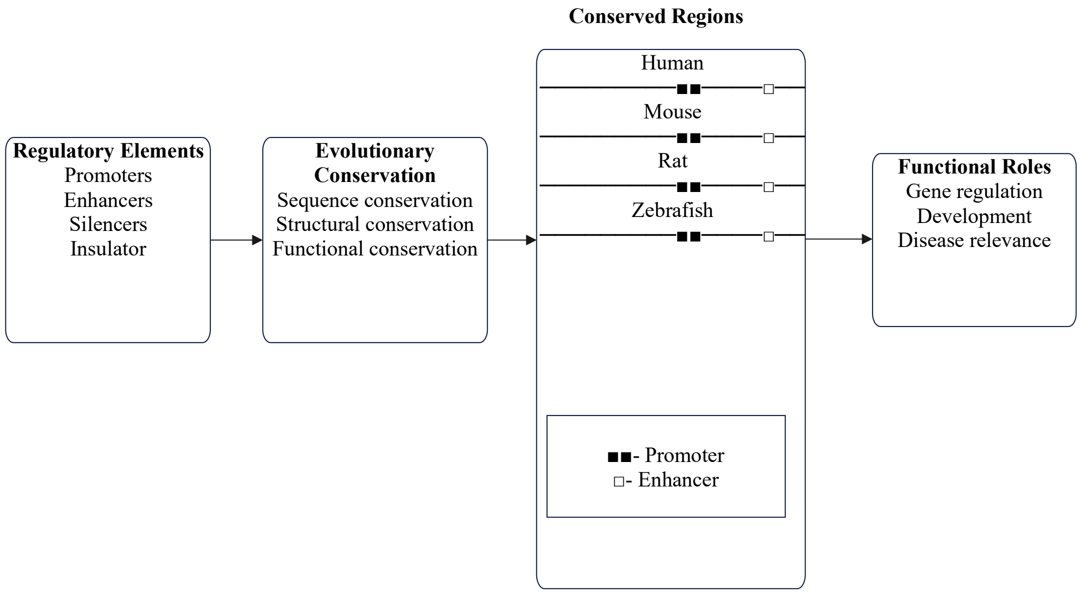

Figure 1. Cross-species conservation of regulatory elements across representative genomes

Figure 1. The conservation of regulatory elements that display parallel conserved promoter regions and enhancer regions in representative species, which implies conservation of gene regulatory sequences during evolution.

4. Comparative Genomic and Functional Characterization Approaches

Comparative genomic analysis is a key factor in the identification of conserved regulatory elements (CREs) in the genomes of multiple species based on sequence similarity and evolutionary conservation. Pairwise and multiple sequence alignment is a popular method of sequence alignment that is used to find conserved non-coding regions that might be interpreted as regulatory elements. The methods facilitating the discovery of the conserved motifs and shared genomic regions across the species allow evidence concerning the regulatory architecture. Moreover, phylogenetic footprinting has become an effective technique of identifying conserved regulatory sequences using orthologues genomic regions of many species hence differentiating between functional and neutral sequences. In addition to the methods based on sequences, the methods of functional genomics have greatly contributed to the identification and characterization of CREs. Chromatin immunoprecipitation sequencing (ChIP-seq) can be used to map the locations of transcription factor binding, and assay of transposase-accessible chromatin by sequencing (ATAC-seq) can be used to understand chromatin accessibility and activity of regulatory factors. These methods allow combination of the epigenomic data with comparative genomics which can enhance the prediction capability of CRE. In Table 2, a summary of the commonly used tools and databases that can be used to support these analyses is presented and illustrated with the applications of these tools in regulatory elements identification and functional annotation. Experimental validation is also an essential measure in verifying the relevance of the elements of regulation that were predicted in terms of functionality. The activity of candidate CREs in vivo and in vitro is commonly evaluated by the use of techniques including CRISPR-based genome editing and reporter assays. These methods enable the researcher to determine the regulatory effects of particular genomic regions and confirm their influence on expression of genes. The general workflow of CRE identification and validation (Figure 2) involves multi-species genomic data (usually) integrated with comparative genomic analysis, candidate regulatory elements identification, and functional and experimental validation (all) eventually resulting in the validation of biologically relevant CREs.

Table 2. Tools and databases for comparative genomic and functional analysis of regulatory elements

Tool/Database	Type	Purpose	Application
UCSC Genome Browser	Database	Genome visualization	Cross-species comparison
ENCODE	Database	Functional genomics data	Regulatory annotation
VISTA Enhancer Browser	Database	Enhancer validation	Experimental validation
HOMER	Software	Motif discovery	TF binding analysis
MEME Suite	Software	Motif identification	Regulatory motif detection
Galaxy Platform	Web tool	Genomic data analysis	Integrated workflows

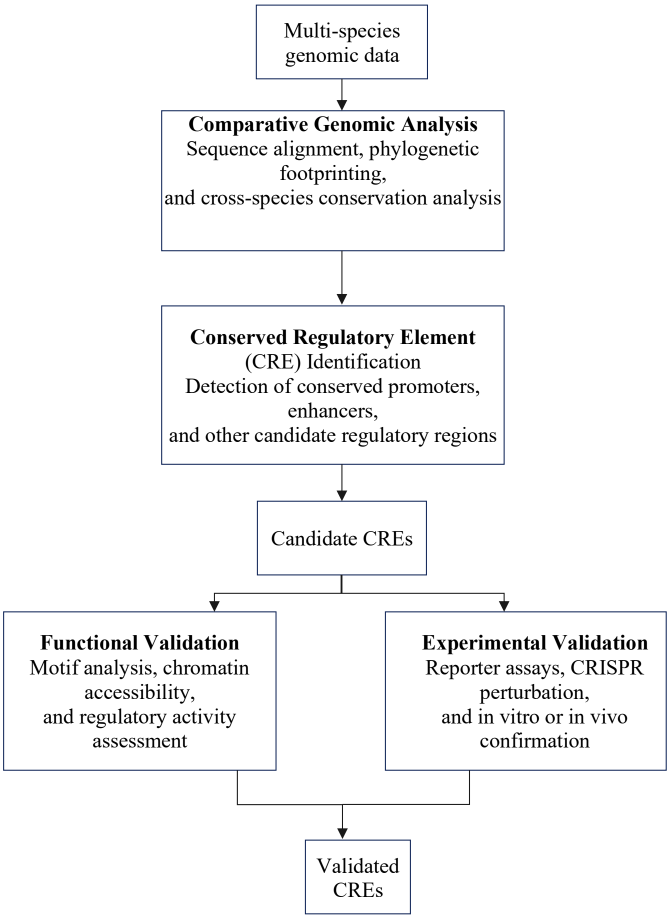

Figure 2. Workflow for analysis and validation of conserved regulatory elements

Figure 2. Conserved regulatory element analysis and validation workflow depicts the chronological approach of integrating research data of multi-species genomes, comparative genomics analysis, CREs identification, candidate selection, and validation of functional and experimental validation to attain validated regulatory elements.

5. Computational Models of Predicting Conserved Regulatory Elements.

The computational methods have become critical in discovery and forecast of the existence of conserved regulatory elements (CREs), especially when dealing with large genomic data sets. The methods that had been used extensively to identify conserved non-coding regions across species were traditional sequence-based methods, which are alignment and phylogenetic footprinting. Nevertheless, these methods usually fail on detecting regulatory factors having weak or non-linear conservation pattern. To solve these problems the sophisticated computational methods have been worked out to make the prediction of CRE more accurate and efficient. Recent computational approaches can use a variety of genomic characteristics, such as sequence motifs, conservation scores, chromatin accessibility, and epigenomic signatures to improve regulatory element detection. The use of machine learning and data-driven models has also improved this area as the complex patterns of genomic data that cannot be easily modelled with traditional approaches could be identified. Through these methods, it is possible to predict regulatory factors even in situations where species have constrained sequence conservation and functional conservation in species. An example of computational prediction of CREs is shown in Figure 3, where a structured workflow is used in which multi-species genomic sequences are used as the input, and then extracted features are made using features based on both sequence and epigenomic features. The features are then analyzed with computational models in order to determine the candidate regulatory elements, which are further analyzed by functional annotation using other genomic and transcriptomic data. This workflow makes it possible to identify a systematic and scalable method of CRE identification in the middle of sequence-based prediction and functional interpretation. All in all, computational methods offer an effective platform to develop conserved regulatory element studies and thus predict them more precisely, analyze inter-species, and multi-dimensionally analyze genomic data.

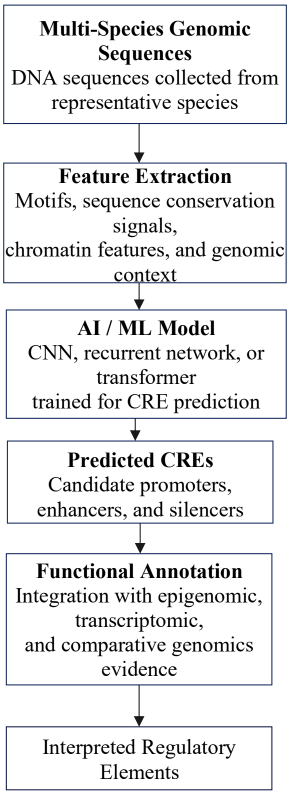

Figure 3. Workflow for computational prediction and annotation of conserved regulatory elements
The figure shows the sequential sequence of the process of detecting conserved regulatory elements (CREs) by using computational methods. Multi-species genomic stream information is used as the workflow starting point with feature extraction that makes use of sequence motifs, conservation signals, and epigenomic characteristics. These are features that are processed with computational models to give candidate regulatory elements, which are further annotated by functional evidence (by integrating with transcriptomic, epigenomic and comparative genomics evidence). The end result is expressed regulatory elements of possible biological interest.

6. Applications and Biological Implications.

The conserved regulatory elements (CREs) are applied extensively in various fields of biology because of their central nature in regulating genes expression, and genomic stability. CREs are important in evolutionary biology because they can give important understanding on underlying processes in species conservation and divergence. The fact that the regulatory sequences are conserved in species that are far apart means that the sequences are critical to ensuring that the biological functions are core. Comparative genomic research has revealed that, although the coding sequences can be relatively stable, the difference in regulatory elements can be used to cause phenotypic diversity and adaptation. CREs are therefore important determiners of evolution processes, as they affect patterns of gene expression without modifying underlying protein-coding sequences. Conserved regulatory elements are also a major issue in the context of disease genomics to gain insight into the genetic basis of different disorders. These non-coding regulatory regions play a role in regulation leading to abnormal expression of genes involved in the occurrence of diseases, including cancer, neurological and abnormal development. The genome-wide association studies (GWAS) have shown that a large number of disease-related variants are found in non-coding regions, especially in conserved regulatory elements. This underscores the need to focus on the study of CREs as a way of identifying possible biomarkers and therapeutic targets. Moreover, the knowledge about the regulatory structure of the genome can help develop precision medicine methods, in which specific interventions may be developed, depending on specific genomic data. The conserved regulatory elements also play an essential role in developmental regulation whereby they regulate the time and space-specific expression of genes in the growth and differentiation of the organism. These factors control major developmental pathways and are in contact with transcriptional factors and epigenetic processes, where appropriate cell fate development and tissue formation are assured. Their role in the conservation of the standard biological processes is proven by the fact that developmental regulatory elements are preserved in different species. Defects in development and congenital disorders are possible in the case of disruption of these aspects, again highlighting their functional importance. On the whole, CREs study is a useful source of information about the intricate regulatory systems that operate in biological systems and promising directions in the evolution, medical, and developmental biology research.

7. CHALLENGES AND LIMITATIONS

Although much has been done in terms of identifying and defining the conserved regulatory elements (CREs), a number of challenges and limitations remain as to impede the entire and precise comprehension of the roles of CREs across species. The unfinished annotation of non-coding regions of most genomes, especially non-model organisms is one of the main problems. Although large genomic resources exist in well-studied species like humans, mice, and zebrafish, a big percentage of the species do not have high quality reference genomes and functional annotations. Such an imbalance produces a bias in comparative genomic studies and restricts the generalization of the results to a wide evolutionary range. This has led to the fact that lots of potentially significant regulatory factors within underrepresented species are unknown or not well studied. The other important constraint is brought about by complex and context-dependent nature of regulatory elements. CREs are highly sensitive to various factors that determine their activity such as the type of cell, developmental stage, the environmental conditions, and epigenetic modifications. Because of this, sequence conservation may not necessarily be a good predictor of functional relevance. Not all of the conserved sequences can be regulated under particular conditions, and the sequences with a low degree of sequence conservation might also have functional purposes. Such a functional plasticity makes comparative genomic analysis more difficult to interpret and urges the need to combine on several layers of biological data. In addition, regulatory factors tend to operate within bigger networks, and it is hard to identify certain regulatory effects to the elements working by themselves. There are also a number of limitations of computational methods to identify CREs which are associated with sensitivity, specificity, and scalability. The commonly used traditional approaches like sequence alignment and phylogenetic footprinting can be good in the detection of highly conserved regions, but may fail to detect regulatory elements that have weak, divergent, or non-linear conservation patterns. Moreover, there is a tendency of inconsistent results of the different computational tools, and as a consequence, there are difficulties in standardization and reproducibility. The combination of the large scale genomic, epigenomic and transcriptomic databases also adds complexities in terms of data heterogeneity, noise and computational resource demands. The analysis and handling of these large volume data sets demand complex algorithms and high speed computing systems and they are not always easily available. Another significant bottleneck is the experimental validation of the elements of the regulatory elements that have been predicted. Reporter assays, chromatin immunoprecipitation, and CRISPR based genome editing techniques offer useful functional information but are resource intensive, time consuming and cannot be scaled to provide genome wide analyses. Moreover, experimental outcomes may be varied with regard to experimental conditions and model systems, which brings in another aspect of variability. All these issues emphasize the importance of more integrative methods that unite the use of computational predictions with high throughput function validation, standardized procedures and cross-species data collections to enhance the accuracy, reliability, and biological significance of conserved regulatory element identification and characterization.

8. Future Perspectives

Identification of conserved regulatory elements (CREs) is also likely to be studied further as the staples of high-throughput sequencing technologies, and integrative methods of computation continue to develop. The integration of multi-omics data, such as genomics, epigenomics, transcriptomics and proteomics, can be considered one of the most promising directions to help understand the regulatory mechanisms in several species more comprehensively. The combination of these different datasets will allow much more precise discovery of functional regulatory elements and clarify the role of these elements in complex biological systems. New technologies like single-cell sequencing have the potential to revolutionize the field as they will allow the study of the activity of regulatory elements at an unprecedented level of detail. These solutions enable researchers to examine cell-type specific and context dependent regulatory processes which is one of the primary limitations of the existing bulk genomic analysis. Moreover, the development of spatial transcriptomics will also contribute to the further insight into the mechanisms of the regulatory components work in the context of particular tissues structure and maturation. The methods of artificial intelligence (AI) and machine learning (ML) are also likely to become the key ones in the future. Deep learning models are capable of searching large genetic datasets to determine complicated trends and forecast the regulatory variables more precisely compared to traditional approaches. Even in those instances where the sequence conservation is loose but the functional conservation maintained, these techniques can be used to aid cross-species extrapolation of regulatory elements. With the ongoing improvement of the computational power and complexity of algorithms, the use of AI-based models will become a part of comparative genomics. Moreover, with the development of genome editing, especially the CRISPR-based systems, it will be possible to perform high-throughput functional validation of predicted CREs. These tools enable the fine manipulation of regulatory regions to give direct evidence on their involvement in the expression of genes which determine phenotyping. It will also be important to expand the research to cover a wider scope of species to involve more non-model species to understand how regulatory factors evolve. In general, the combination of high-level experimental procedures, computer modeling, and cross-species studies should result in a breakthrough within the field and the subsequent progression to the comprehension of the gene regulation and its consequences in evolution, development, and disease.

9. CONCLUSION

The concept of conserved regulatory elements (CREs) is an essential element of the genomic regulatory environment as it has a pivotal role in regulating the expression of genes in a variety of species. Comparative genomic strategies have resulted in characterizing and determining these elements as well as their importance in sustaining the key biological functions. Their conservation shows the evolutionary importance of the promoters, the enhancers, and the additional regulatory regions and emphasizes their role in developmental processes and their contribution in cellular differentiation, and in maintaining the phenotype. These components are not only able to maintain fundamental biological processes but they also offer evidence on the role of regulatory disparities that contribute to the phenotypic divergence among species. This review has given an in-depth account of the CREs, the classification, evolutionary history, functional roles and methods used to identify them and validate them. The sensitivity of the identification and analysis of conserved regulatory regions has been greatly improved by the development of computational methods, like sequence alignment and phylogenetic footprint, as well as the high-throughput functional genomics methods. Moreover, experimental validation techniques have also been used in establishing the biological relevance of the elements predicted which have broken the gap between computational prediction and functional insight. Inclusion of computational strategies adds more stress to the increasing significance of data-driven solutions in enhancing the prediction and interpretation of CRE, specifically, in large-scale genomic studies. Although these have been achieved, some of the challenges still exist such as poor annotation of genomes, activation of regulatory activity in specific situations, and the inability to test large functional validations. These aspects demonstrate that the regulatory systems are complex, and more powerful and integrative frameworks of analysis should be taken into consideration. These challenges will demand an integrated approach that will involve integration of multi-omics data, better computational modeling and scalable experimental validation. It will also be necessary to expand the research beyond the model organisms in order to gain a more comprehensive view on regulatory conservation and diversity. New technologies like state-of-the-art computational modeling, single-cell genomics, and CRISPR-based genome editing have a lot of potential to overcome the limitations and move the field forward. Computer-based and data-driven methods will lead to better discovery of subtle regulatory patterns and prediction across species of functional elements. With these developments, it will be possible to more accurately characterize CREs and apply them in a variety of biological and biomedical settings. Conclusively, conserved regulatory elements are a fast-growing area of studies with expansive implications in evolutionary biology, disease genomics and in developmental regulation. The further study of CREs will not only improve our comprehension of the role of gene regulation but it will also lead to new developments in precision medicine, biotechnology and functional genomics, and end up applying the findings of genomic studies to practical uses.

REFERENCES

1. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., ... & Sandelin, A. (2014). An atlas of active enhancers across human cell types and tissues. *Nature, 507*(7493), 455–461.
2. Andersson, R., & Sandelin, A. (2020). Determinants of enhancer and promoter activities of regulatory elements. *Nature Reviews Genetics, 21*(2), 71–87.
3. Arnold, C. D., Gerlach, D., Stelzer, C., Boryń, Ł. M., Rath, M., & Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science, 339*(6123), 1074–1077.
4. Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Current Protocols in Molecular Biology, 109*(1), 21–29.
5. Ernst, J., & Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods, 9*(3), 215–216.
6. Gusmao, E. G., Allhoff, M., Zenke, M., & Costa, I. G. (2016). Analysis of computational footprinting methods for DNase sequencing experiments. *Nature Methods, 13*(4), 303–309.
7. Heinz, S., Romanoski, C. E., Benner, C., & Glass, C. K. (2015). The selection and function of cell type-specific enhancers. *Nature Reviews Molecular Cell Biology, 16*(3), 144–154.
8. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Kheradpour, P., ... & Roadmap Epigenomics Consortium. (2015). Integrative analysis of 111 reference human epigenomes. *Nature, 518*(7539), 317.
9. Kvon, E. Z. (2015). Using transgenic reporter assays to functionally characterize enhancers in animals. *Genomics, 106*(3), 185–192.
10. Long, H. K., Prescott, S. L., & Wysocka, J. (2016). Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell, 167*(5), 1170–1187.
11. Lowe, C. B., Kellis, M., Siepel, A., Raney, B. J., Clamp, M., Salama, S. R., ... & Haussler, D. (2011). Three periods of regulatory innovation during vertebrate evolution. *Science, 333*(6045), 1019–1024.

12. Nord, A. S., Blow, M. J., Attanasio, C., Akiyama, J. A., Holt, A., Hosseini, R., ... & Visel, A. (2013). Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell, 155*(7), 1521–1531.
13. Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., & Bejerano, G. (2013). Enhancers: five essential questions. *Nature Reviews Genetics, 14*(4), 288–295.
14. Rada-Iglesias, A., Bajpai, R., Prescott, S., Brugmann, S. A., Swigut, T., & Wysocka, J. (2012). Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell, 11*(5), 633–648.
15. Schmidt, D., Wilson, M. D., Ballester, B., Schwalie, P. C., Brown, G. D., Marshall, A., ... & Odom, D. T. (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science, 328*(5981), 1036–1040.
16. Seemann, S. E., Mirza, A. H., Hansen, C., Bang-Berthelsen, C. H., Garde, C., Christensen-Dalsgaard, M., ... & Gorodkin, J. (2017). The identification and functional annotation of RNA structures conserved in vertebrates. *Genome Research, 27*(8), 1371–1383.
17. Shlyueva, D., Stampfel, G., & Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nature Reviews Genetics, 15*(4), 272–286.
18. Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., ... & Stamatoyannopoulos, J. A. (2012). The accessible chromatin landscape of the human genome. *Nature, 489*(7414), 75–82.
19. Villar, D., Berthelot, C., Aldridge, S., Rayner, T. F., Lukk, M., Pignatelli, M., ... & Odom, D. T. (2015). Enhancer evolution across 20 mammalian species. *Cell, 160*(3), 554–566.
20. Zerbino, D. R., Wilder, S. P., Johnson, N., Juettemann, T., & Flicek, P. R. (2015). The Ensembl Regulatory Build. *Genome Biology, 16*(1), 56.