Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Toxicology

Online resources for studies of genome biology and epigenetics

Paul J. Lee[a,1], Mayank NK Choudhary[b,1] and Ting Wang[b]

Abstract

Environmental exposure to chemical toxins alters epigenetic modifications that culminate in altered cellular gene expression without changing the underlying DNA sequence. The complex interplay between the layers of epigenetic regulators ultimately results in observed cellular phenotype. This review highlights epigenetics annotations assayed in the Encyclopedia of DNA Elements (ENCODE) community resource project—a publically accessible database for understanding genomic function, development and disease etiologies. We outline the multiple levels of epigenetic control (DNA methylation, chromatin accessibility, histone modifications, genome topology) with their associated interrogation methodology. We explore the limitations and strengths of each methodology at every epigenetic checkpoint. This review points readers to epigenetic resources that have gathered focused scientific data and directs them toward data visualization tools that can help answer questions related to epigenetic controls. The purpose of this review is to highlight online resources available to toxicological epigenetic researchers that can help fast track novel insights using already curated reference epigenome datasets.

Addresses

[a] Department of Pathology and Immunology, Washington University, St. Louis, MO, USA

[b] The Edison Family Center for Genome Sciences & Systems Biology, Department of Genetics, Washington University, St. Louis, MO, USA

Corresponding author: Department of Genetics, Washington University School of Medicine, 4515 McKinley Avenue, Campus Box 8510, St. Louis, MO 63110, USA. Wang, Ting (twang@genetics.wustl.edu)

[1] These authors contributed equally to this work.

Current Opinion in Toxicology 2017, **6**:34–41

This review comes from a themed issue on **Genomic Toxicology: Epigenetics**

Available online 8 August 2017

For a complete overview see the Issue and the Editorial

http://dx.doi.org/10.1016/j.cotox.2017.07.007

Keywords

Next generation sequencing, Online database, Epigenome, Methylome, Methylation, Chromatin accessibility, Histone modifications, 3D genome, TADs, ENCODE, ROADMAP.

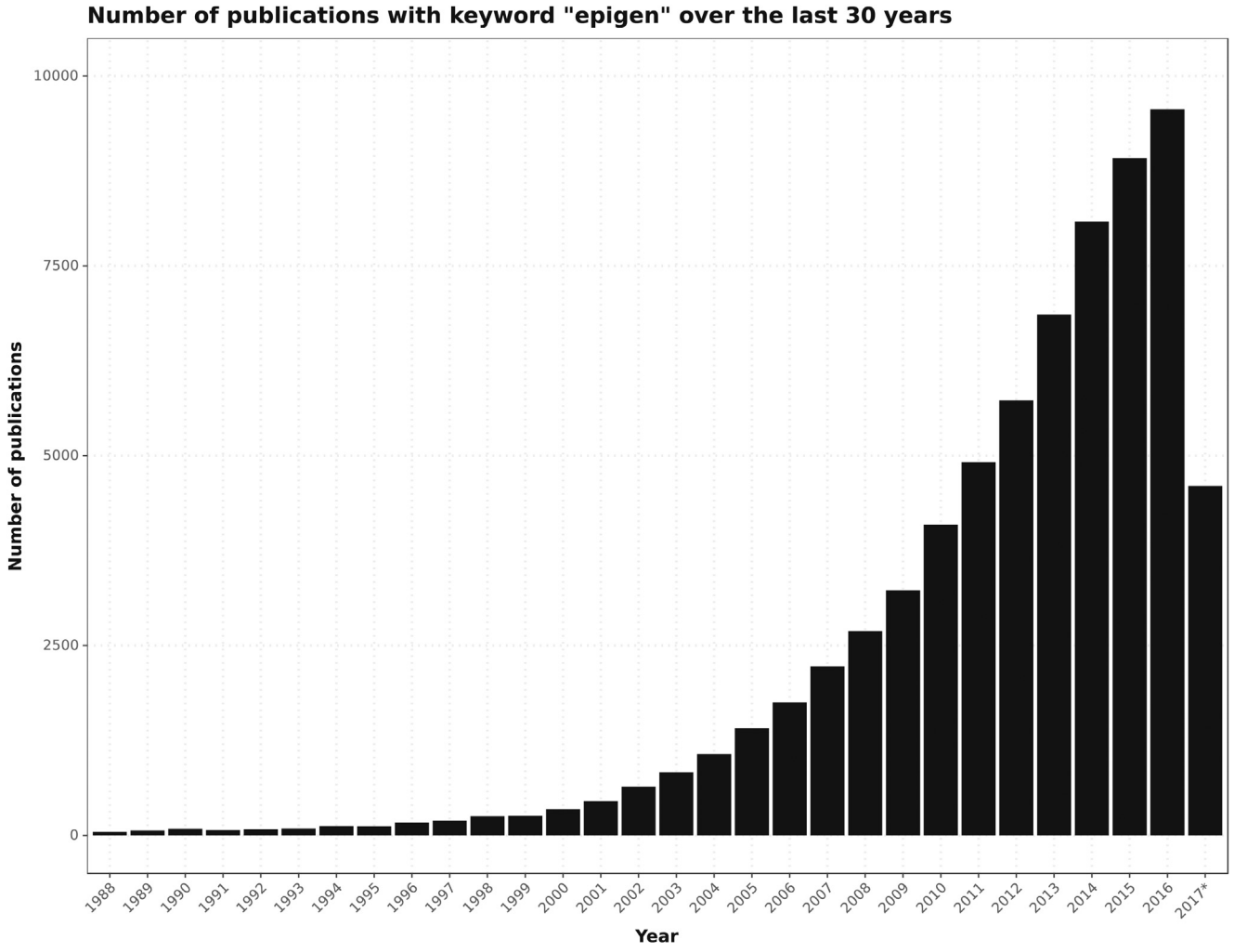
1. Introduction

Recent advances in sequencing and bioinformatics has sparked interest in the research community (Fig. 1) and accelerated our insights into fundamental biological questions regarding the epigenome [1,2]. The Human Genome Project (HGP)—completed in 2003—gave us the first reference genome with an estimated 19,000–20,000 human protein coding genes and approximately 2.9 billion base-pair sequence that makes up the human genome [3]. The $3 billion HGP is estimated to have a 178× return on investment for every public dollar invested in the original sequencing [4]. Studies correlating sequence variants to common multifactorial diseases and traits such as heart disease, diabetes, and obesity have indicated risks unattributable to a single genetic cause, but instead to the impact of lifestyle and environmental factors [5]. Analysis and comparison of the epigenomes from healthy and diseased cells that arise because of aging, exposure to environment, toxins, etc. are essential for understanding the drivers of these complex diseases. Hence, a repository of comprehensive catalogs of public genomic resources is a prerequisite in achieving the goal of precision medicine.

The generation of enormous amounts of sequencing data has driven the need for advanced storage strategies that facilitate shared access—an open-access biomedical repository [6]. The benefits of using one include the reinterpretation potential available for researchers, which often leads to novel insights from pre-existing data [7]. These databases should be centralized, easily accessible, and allow for bulk data download capabilities. Having a variety of epigenomic datasets in a single repository allows for integrated and comprehensive data analysis [8]. On the contrary, disparate data in separate information silos diminishes data value. It is this orthogonal genomic/epigenomic data integration that maximizes the insights from large bioinformatics data—the whole being greater than the sum of its parts [9].

A large-scale effort in generating reference epigenomic annotations is the Encyclopedia of DNA Elements (ENCODE) Consortium [10]. It is an international collaboration of diverse research groups supported by the National Human Genome Research Institute (NHGRI) to identify all the functional elements in the genome including regulatory elements that control cell and gene activity in humans and other mammals [11]. ENCODE is a *community resource project* that aims to assemble a comprehensive encyclopedia of functional elements, describing their identity and precise location in a variety of genomes. The epigenomic data sets, regulatory annotations and integrative analyses have resulted in the most comprehensive maps of the

Fig. 1



An increase in publications related to epigenetics is observed in the last few decades. PubMed (http://www.ncbi.nlm.nih.gov/pubmed) search was performed in May 2017 for publications that mention the wildcard epigen* (to capture epigenetics, epigenomics, etc.). With the bulk of 2017 remaining, it is projected to rise over to 11,000 publication mentions.

epigenomic landscape so far across the largest collection of primary cells and tissues. The ENCODE map is of broad use to the scientific and biomedical communities, for studies of genome interpretation, gene regulation, cellular development and differentiation, genome evolution, genetic variation and human disease.

Epigenetic alterations play an important role in mediating cellular processes in response to environmental exposures (pesticides, toxins, mutagens, etc.). The chemical and oxidative stresses trigger chemical moiety modifications to the genomic DNA and its associated histone proteins. While the plasticity of the epigenome allows for subtle adaptations to environmental alterations, it predisposes the cell to epigenomic mishaps, which result in altered gene expression and ultimately aberrant cellular phenotype. In this review, we examine how online databases can help investigators answer questions associated with toxicological epigenetics (Fig. 2).

2. ENCODE epigenomic mapping landscape

The genome-wide mapping of the epigenetic marks occurs at many different control levels. The systematic profiling of these different control levels is needed to create a unified understanding of developmental and disease processes [12]. All the methods/assays discussed in this review are freely available on the ENCODE website at: https://www.encodeproject.org/

2.1. DNA methylation

DNA methylation most commonly takes place at the 5th position of cytosine (5-mC) nucleotide. It is involved in