



ELSEVIER

Contents lists available at ScienceDirect

## Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)



### Data Article

# Gene regulation by long purine tracks in brain related diseases



Himanshu Narayan Singh<sup>a,b</sup>, Moganty R. Rajeswari<sup>a,\*</sup>

<sup>a</sup> Department of Biochemistry, All India Institute of Medical Sciences, New Delhi 110029, India

<sup>b</sup> School of Sciences, Noida International University, Gautam Budh Nagar, 203201 Uttar Pradesh, India

#### ARTICLE INFO

##### Article history:

Received 2 August 2015

Received in revised form

15 August 2015

Accepted 24 August 2015

Available online 4 September 2015

##### Keywords:

Purine repeat

Human genome

Gene regulation

Brain disease

#### ABSTRACT

Purine repeats are randomly distributed in the human genome, however, they show potential role in the transcriptional deregulation of genes. Presence of long tracks of purine repeats in the genome can disturb its integrity and interfere with the cellular behavior by introducing mutations and/or triple stranded structure formation in DNA. Our data revealed interesting finding that a majority of genes carrying purine repeats, of length  $n \geq 200$ , were down regulated and found to be linked with several brain related diseases [1]. The unique feature of the purine repeats found in the present study clearly manifests their significant application in developing therapeutics for neurological diseases.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

#### Specifications table

Subject area	Biology
More specific subject area	Genetics, Bioinformatics
Type of data	Table, Software generated files

DOI of original article: <http://dx.doi.org/10.1016/j.gene.2015.07.007>

\* Corresponding author.

E-mail address: [rajeswari3011@hotmail.com](mailto:rajeswari3011@hotmail.com) (M.R. Rajeswari).

<http://dx.doi.org/10.1016/j.dib.2015.08.024>

2352-3409/© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

How data was acquired	Software generated
Data format	Analyzed
Experimental factors	Purine repeats ( $n \geq 200$ ) were searched in the human genome and also tried to explore their association with neurological disorders.
Experimental features	Purine repeat were searched by the help of home-made PERL script and further mapped them with neurological disorders
Data source location	New Delhi, India
Data accessibility	Data is supplied in this article

---

### Value of the data

---

- Identified purine repeats (PR,  $n \geq 200$ ) are unique in the human genome. Therefore, genes carrying purine repeats can be used as potential therapeutic tools in controlling gene expression and also in sequence-specific drug delivery.
  - The data will be helpful to explore the risk associated with acquiring disease causing mutations related to diseases.
  - The data will also be useful to study of evolutionary dynamics.
- 

## 1. Data, experimental design, materials and methods

### 1.1. Data resources

In present study, four data resources were utilized viz. (i) Human Genome Sequence: NCBI/Genome database; (ii) gene annotation: Ensemble Genome Browser; (iii) gene-disease association: GenAtlas database; and (iv) expression datasets: NCBI/GEO database. [Table 1](#)

## 2. Algorithm developed for purine repeat search

An indigenous PERL script “PuRepeatFinder.pl” was developed to locate PRs,  $n \geq 200$ , in the human genome. The tool enlists the PRs in the chronological order of its genomic coordinates along with PR-length and sequence. The script implements the knowledge based window-shift algorithm, and identify only uninterrupted, non-overlapping purine repeats.

## 3. Web tools

We have utilized two web-tools: (i) non-B DNA Motif Search Tool (nBMST): to search for the mirror repeat motifs with the identified PRs. It searches for the perfect and imperfect mirror repeats within the provided sequences [2] and (ii) Idiographica: to show the distribution of PR-genes on the chromosomes [3].

## 4. Microarray data analysis

Two open source R-packages of Bioconductor project viz. limma: used for agilent based microarray data, and affy: for affymatrix based microarray data, were used to calculate gene expression levels. Expression computation involves three steps: (i) background correction, (ii) normalization and (iii)

Download English Version:

<https://daneshyari.com/en/article/174877>

Download Persian Version:

<https://daneshyari.com/article/174877>

[Daneshyari.com](https://daneshyari.com)