Key Points:

1. Database backend
2. Schema-free
3. Choice : MongoDB, Apache Drill as a backend
4. Using R as an interface to query data from these databases. Extract the data as GRanges to make it easy to use for network approaches
5. Benefits : fast, free of cost

In order to efficiently exploit the disease and trait-associated genetic variants identified by the genome-wide association studies and efficiently use the results to develop a transcriptional regulatory network, it is necessary to construct a genome-wide regulatory database.This will help downscale the problems of data deluge from GWAS encompassing millions of patients by integrating information, thereby making it easier to store, retrieve, collect data.

The database backend architecture comprises MongoDB and Apache Drill to store eQTL, annotation and transcription factor data. TxRegQuery is an R package that provides users with tools for information access and retrieval from the database. TxRegQuery makes it possible to extract data in the form of a GRange object, which could be used for building a transcriptional regulatory network downstream. Use of mongoDb and apache drill would help scientists perform deep queries in a very fast, cost free efficient manner.

GWAS are growing at an increasing rate to encompass millions of patients leading to a deluge of genetic association data. To efficiently exploit and use the disease and trait-associated genetic variants identified, it is beneficial to build genome-wide regulatory networks.