*Published online 25 August 2021 Nucleic Acids Research, 2022, Vol. 50, Database issue* ***D39–D45***

*https://doi.org/10.1093/nar/gkab740*

3raQTL-atlas: an atlas of 3rUTR alternative polyadenylation quantitative trait loci across human normal tissues

**Ya Cui 1, Fanglue Peng 2, Dan Wang3, Yumei Li1, Jason Sheng Li1, Lei Li1,\* and Wei Li 1,\***

1Division of Computational Biomedicine, Department of Biological Chemistry, School of Medicine, University of California, Irvine, Irvine, CA 92697, USA, 2Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030, USA and 3Department of Medicine, Division of Cardiology, University of California, Los Angeles, Los Angeles, CA, 90095, USA

Received July 31, 2021; Revised August 10, 2021; Editorial Decision August 11, 2021; Accepted August 13, 2021

# ABSTRACT

**Genome-wide association studies (GWAS) have identified thousands of non-coding single-nucleotide polymorphisms (SNPs) associated with human traits and diseases. However, functional interpretation of these SNPs remains a significant challenge. Our**

# recent study established the concept of 3r un- translated region (3rUTR) alternative polyadenylation (APA) quantitative trait loci (3raQTLs), which can

**∼**

**be used to interpret 16.1% of GWAS SNPs and**

# are distinct from gene expression QTLs and splic- ing QTLs. Despite the growing interest in 3raQTLs, there is no comprehensive database for users to

**search and visualize them across human normal** [**tissues. In the 3raQTL-atlas (https://wlcb.oit.uci.edu/ 3aQTLatlas), we provide a comprehensive list of**](https://wlcb.oit.uci.edu/3aQTLatlas) **3raQTLs containing 1.49 million SNPs associated with APA of target genes, based on 15,201 RNA-seq**

**∼**

# samples across 49 human Genotype-Tissue Expres- sion (GTEx v8) tissues isolated from 838 individu-

**als. The 3raQTL-atlas provides a 2-fold increase in**

**∼**

# sample size compared with our published study. It also includes 3raQTL searches by Gene*/*SNP across tissues, a 3raQTL genome browser, 3raQTL boxplots, and GWAS-3raQTL colocalization event visualization. The 3raQTL-atlas aims to establish APA as an emerg-

**ing molecular phenotype to explain a large fraction**

# of GWAS risk SNPs, leading to significant novel in- sights into the genetic basis of APA and APA-linked susceptibility genes in human traits and diseases.

**INTRODUCTION**

Genome-wide association studies (GWAS) have identified

*>*100 000 single-nucleotide polymorphisms (SNPs) associ- ated with complex traits and diseases in humans. However, the functional interpretation the effects of these SNPs is difficult, as the SNPs often lie in non-coding regions and do not directly affect protein-coding regions. To bridge the gap between GWAS non-coding variants and human phe- notypes, a quantitative trait locus (QTL)-based analysis has been used to evaluate the effects on various molecular phe- notypes. In particular, gene expression (eQTL) and splic- ing (sQTL) analyses have successfully explained the target genes and functional mechanisms of numerous GWAS risk loci ([1–4](#_bookmark12)). Despite massive efforts on eQTLs and sQTLs, a large fraction of GWAS risk loci remains unexplained ([5](#_bookmark15),[6](#_bookmark18)). Alternative polyadenylation (APA), which occurs in

*>*70% of human genes, is a major mechanism of post- transcriptional regulation under diverse biological condi- tions, tissues and cell types ([7–10](#_bookmark21)). By changing the posi- tion of polyadenylation sites, APA can generate transcripts

with either long or short 3r untranslated regions (3rUTRs) that contain different cis-regulatory elements, such as bind-

ing sites of RNA-binding proteins or miRNAs. This leads to altered translation efficiency, cellular localization, and stability of transcripts ([9](#_bookmark23),[11](#_bookmark2)) independent of gene expres- sion or splicing. Disruption of the key APA regulators (e.g. *PABPN1*, *CDK12* and *NUDT21*) leading to global APA changes has been linked to serious human diseases, including oculopharyngeal muscular dystrophy ([12](#_bookmark3)), neu- ropsychiatric disease ([13](#_bookmark4)), leukemia ([14](#_bookmark5)), and glioblastoma ([15](#_bookmark6)).

In addition, mounting evidence suggests that genetic variations that affect APA usage in individual genes can confer the risks of many diseases, including Parkinson’s dis- ease ([16](#_bookmark7)), systemic lupus erythematosus ([17](#_bookmark8),[18](#_bookmark9)), multiple cancer types ([19](#_bookmark10)), and diabetes ([20](#_bookmark11),[21](#_bookmark14)). For example, a com-

Downloaded from https://academic.oup.com/nar/article/50/D1/D39/6357732 by Baylor College of Medicine user on 01 September 2022

\*To whom correspondence should be addressed. Tel: +1 949 824 6567; Email: [we](mailto:wei.li@uci.edu)[i.li@uci.edu](mailto:i.li@uci.edu) Correspondence may also be addressed to Lei Li. Email: [lei.li.bioinfo@gmail.com](mailto:lei.li.bioinfo@gmail.com)

ⓍC The Author(s) 2021. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

[(http:*//*creati](http://creativecommons.org/licenses/by-nc/4.0/))v[ecommons.org*/*licenses*/*by-nc*/*4.0*/*),](http://creativecommons.org/licenses/by-nc/4.0/)) which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

***D40*** *Nucleic Acids Research, 2022, Vol. 50, Database issue*

Downloaded from https://academic.oup.com/nar/article/50/D1/D39/6357732 by Baylor College of Medicine user on 01 September 2022

mon variant (rs356165) at the 3rUTR of *SNCA* (coding α- synuclein protein) can increase *SNCA* long 3rUTR usage, which enhances the accumulation of α-synuclein protein

in mitochondria and further contributes to a high risk of Parkinson’s disease ([16](#_bookmark7)). Similarly, rs10954213 at the *IRF5*

3rUTR locus can shorten the 3rUTR of *IRF5*, which al- ters mRNA stability and further results in high systemic

lupus erythematosus susceptibility ([17](#_bookmark8)). Taking the advan- tage of large-scale transcriptome and genotype data from

the Genotype-Tissue Expression (GTEx) project (version 7), our recent study established the concept of 3rUTR APA quantitative trait loci (3raQTLs), which can be used to in-

∼

terpret 16.1% of GWAS SNPs and are largely distinct from eQTLs and sQTLs ([22](#_bookmark16)). 3raQTLs provide consequen- tial supplements to the functional interpretation of non- coding variants. Despite the growing interest in 3raQTLs, there is no comprehensive database for users to search and visualize these 3raQTLs across human normal tissues.

We developed a database for 3raQTLs, termed 3raQTL-

atlas. 3raQTL-atlas contains 1.49 million SNPs associ- ated with the APA of target genes, based on 15,201 RNA-

∼

seq samples across 49 human GTEx (version 8) tissues iso- lated from 838 individuals. The 3raQTL-atlas not only pro- vides a 2-fold increase in sample size compared with our published study ([22](#_bookmark16)) but also includes 3raQTL searches

∼

by Gene*/*SNP across tissues, a 3raQTL genome browser,

3raQTL boxplots, and GWAS-3raQTL colocalization event visualization. The 3raQTL-atlas aims to establish APA as an emerging and important molecular phenotype to explain a

large fraction of GWAS risk SNPs, leading to significant novel biological insights into the genetic basis of APA and APA-linked susceptibility genes in a wide spectrum of hu- man traits and diseases.

# DATA COLLECTION AND PROCESSING

**GTEx data collection and processing**

We downloaded the RNA-seq BAM files of 17 382 human normal samples across 54 tissues in 948 individuals from the GTEx project (dbGaP, phs000424.v8.p2) ([2](#_bookmark13)). The orig- inal RNA-seq reads were aligned to the human genome (hg38*/*GRCh38) using STAR v2.5.3a ([23](#_bookmark17)), with the align- ment parameters described in the GTEx study ([2](#_bookmark13)). We re- moved the BAM files that were either generated from dis- eased tissues and or tissue types with small sample sizes. We also removed RNA-seq BAM files from individuals without genotype data, which were not included in the GTEx analysis freeze. The remaining BAM files were then sorted and converted into bedGraph format using bed- tools v2.17.0 ([24](#_bookmark19)). Genotype data from the GTEx v8 re- lease were called from whole-genome sequencing (WGS) data ([2](#_bookmark13)). Briefly, WGS reads were aligned to the human genome (hg38*/*GRCh38) with BWA ([25](#_bookmark20)). Variants in the variant call format (VCF) file were called using GATK Hap- lotypeCaller v3.5 ([26](#_bookmark22)). After filtering low-quality samples by the GTEx Consortium, the final analysis freeze set con- tained variants called from 838 donors. The final variants were imputed and phased using SHAPEIT v2 ([27](#_bookmark24)). The as- sociated sample description files were downloaded from the GTEx Portal ([www.gtexportal.org](http://www.gtexportal.org/)).

# Quantification of APA usage using DaPars2

Our previously developed DaPars2 ([28](#_bookmark25)) was used to calcu- late relative APA usage, which was measured by the per- centage of distal polyA site usage index (PDUI), from stan- dard RNA-seq data. DaPars2 applies a two-normal mixture model to allow for the joint quantification of APA usage for multiple samples ([28](#_bookmark25)). Using the University of Califor- nia Santa Cruz (UCSC) Table Browser ([29](#_bookmark26)), we first down- loaded the human genes annotation file (hg38*/*GRCh38). Then, the script ‘DaPars Extract Anno.py’ was used to ex-

tract a 3rUTR annotation for each transcript. Afterwards,

we calculated the sequencing depth for each sample, which

was used as an input of DaPars2 for normalizing the se- quencing depth difference of samples, by SAMtools v1.9 ([25](#_bookmark20)). Finally, multiple RNA-seq samples were jointly ana- lyzed to identify de novo APA sites and to calculate the APA usage of each transcript in each sample with DaPars2 ([28](#_bookmark25)).

# 3raQTL mapping for each tissue

3raQTL analysis was performed separately for each tissue, using the genotype and normalized PDUI values as pre-

viously described (Figure [1](#_bookmark0)A) ([22](#_bookmark16)). Briefly, we split the VCF data for each tissue with BCFtools ([25](#_bookmark20)) and further transformed them into genotype matrix files using BioAl- cidae v.2.27.1 ([30](#_bookmark27)). Only variants with a minor allele fre- quency 0.01 were included in further analyses. For each tissue type, we used linear regression by MatrixEQTL ([31](#_bookmark28)) to test the association between normalized PDUI values

≥

and SNPs within a 1-Mb interval of the 3rUTR region. Both known covariates (e.g. sex, RNA integrity number [RIN],

platform, and top five genotype principal components) and unobserved covariates calculated by PEER ([32](#_bookmark29)) were in- cluded in the analysis. The number of PEER covariates for each tissue was selected based on suggestions from the GTEx Consortium: 15, 30 and 35 PEER factors were cho- sen for tissue sample sizes of *<*150, 150–250 and *>*250, re- spectively ([1](#_bookmark12)). By randomly sampling individual labels, 1000 rounds of permutation analysis were conducted to obtain empirical *P*-values for each APA gene. Then, these *P*-values were adjusted with the R package qvalue v.2.0.0 ([33](#_bookmark30)).

# Identification of GWAS-associated 3raQTLs

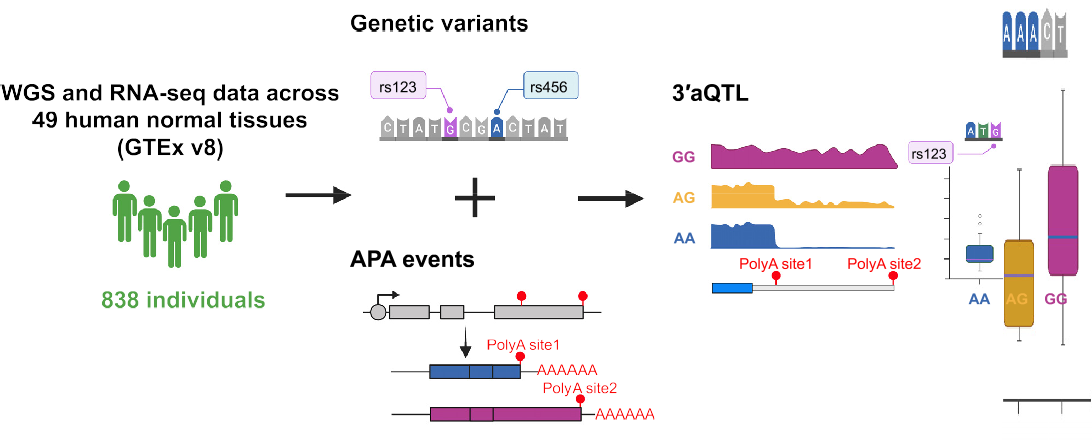
To identify human diseases and traits associated with SNPs, we obtained GWAS risk SNPs (referred to as tag SNPs) from the National Human Genome Research Institute GWAS catalog ([34](#_bookmark32)) (accessed 1 June 2021). SNPs with no dbSNP accessions were removed. Previous studies have sug- gested that causal variants are often not the tag SNPs them- selves, but the SNPs in linkage disequilibrium (LD) with tag SNPs ([35](#_bookmark34),[36](#_bookmark36)). Thus, we extracted a list of SNPs that were in strong LD of European ancestry with the GWAS catalog tag SNP (referred to as LD SNPs). Using the LD cutoff of *r*2 0.8, we finally identified 1 711 210 LD

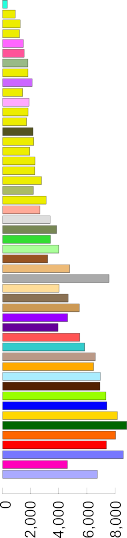
≥

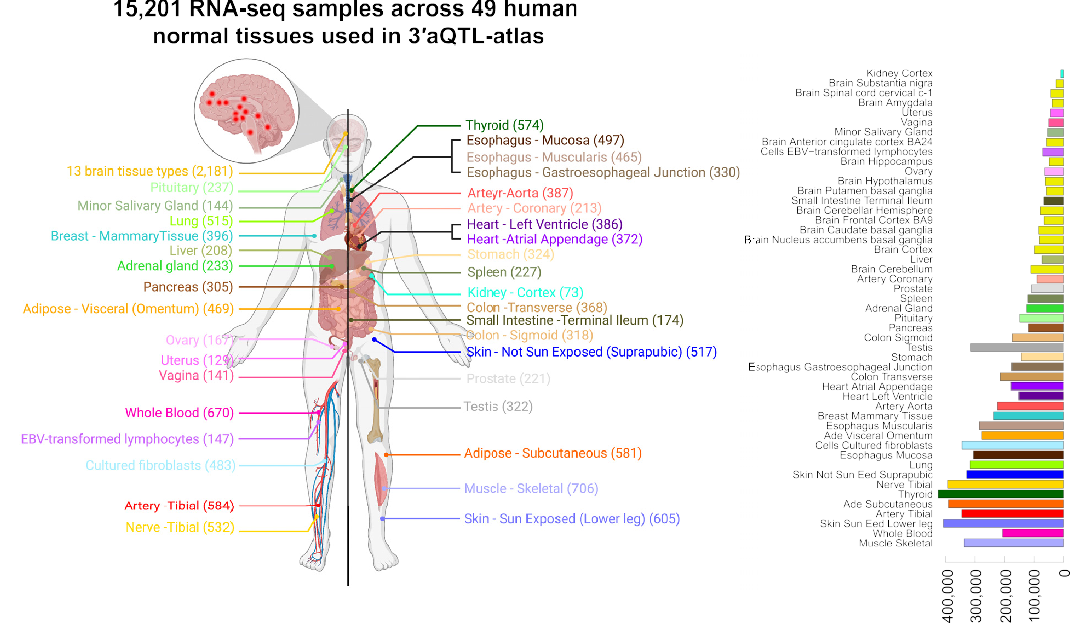
SNPs and tag SNPs, which we refer to as GWAS-associated SNPs. GWAS-associated 3raQTLs were defined when the lead 3raQTLs were overlapped with these GWAS-associated

SNPs.

*Nucleic Acids Research, 2022, Vol. 50, Database issue* ***D41***

**A**

**B C**



**Figure 1.** Data processing and data statistics in 3raQTL-atlas. (**A**) Schematic of overall data processing in the 3raQTL-atlas. (**B**) Distribution of the number of RNA-seq samples for each tissue used in the 3raQTL-atlas. (**C**) Distribution of the number of APA events and significant 3raQTL SNPs (FDR 0.05) for each tissue, sorted by the tissue sample sizes. Each color code indicates a tissue of origin. APA, alternative polyadenylation; WGS, whole-genome sequencing; 3raQTL, 3r untranslated region APA quantitative trait loci.

≤

# Construction of the 3raQTL-atlas website

Downloaded from https://academic.oup.com/nar/article/50/D1/D39/6357732 by Baylor College of Medicine user on 01 September 2022

The 3raQTL-atlas website was constructed on a Linux- based Apache Web server ([http://www.apache.org](http://www.apache.org/)). All pro- cessed and annotated data in the 3raQTL-atlas were stored in MySQL. The R package LocusCompare ([37](#_bookmark27)) and in-

house R scripts were used to perform data analyses and data plotting. The interactive web pages were implemented using HTML, CSS, JavaScript and PHP languages, with several JavaScript libraries (JQuery.js, DataTable.js, NG-Circos.js

and IGV.js) and Bootstrap framework (a popular frame- work for developing interactive websites). The 3raQTL-atlas is freely available and does not require registration or login

for access.

# RESULTS

**Sample summary and 3raQTL landscape of human tissues**

In this version of the 3raQTL-atlas, we analyzed 15,201 RNA-seq samples across 49 human normal tissues from

GTEx version 8 (Figure [1](#_bookmark0)A, B). The RNA-seq sample sizes for each tissue ranged from 73 in the kidney cortex to 706 in skeletal muscle, with a median of 310 (Figure [1](#_bookmark0)B). With the FDR *<*0.05, a total of 1.49 million common genetic

variants associated with 3raQTLs were identified; the me- dian was 30 427 variants per tissue type, with a minimum of

9938 in the kidney cortex and a maximum of 424 628 in thy- roid (Figure [1](#_bookmark0)C). The number of 3raQTL APA events was highly correlated (Pearson correlation *P*-value *<* 2.2e−16,

***D42*** *Nucleic Acids Research, 2022, Vol. 50, Database issue*

Downloaded from https://academic.oup.com/nar/article/50/D1/D39/6357732 by Baylor College of Medicine user on 01 September 2022

*r* 0.91) with the sample size in each tissue (Figure [1](#_bookmark0)C

=

and Supplementary Figure S1). The strong correlation be- tween 3raQTL number and sample size suggests that more 3raQTLs will continue to be identified as additional RNA-

seq datasets become available.

# Data searching, browsing, and visualizing by four modules

[We developed a user-friendly website (3raQTL-atlas; https:](https://wlcb.oit.uci.edu/3aQTLatlas)

[//wlcb.oit.uci.edu/3aQTLatlas) for searching, browsing, and](https://wlcb.oit.uci.edu/3aQTLatlas)

visualizing 1.49 million common genetic variants associ- ated with 3raQTLs across 49 human normal tissues. The 3raQTL-atlas consists of four modules (Figure [2](#_bookmark1)A): a 3raQTL search by Gene*/*SNP (Figure [2](#_bookmark1)B), 3raQTL genome

browser (Figure [2](#_bookmark1)C), 3raQTL boxplot (Figure [2](#_bookmark1)D), and GWAS-3raQTL colocalization event visualization (Figure [2](#_bookmark1)E). In addition, a list of GWAS-associated 3raQTLs are

also provided for users to deeply investigate the mechanisms of 3raQTLs in human traits and diseases.

In the ‘3raQTL search by Gene*/*SNP’ module, users can

search the 3raQTLs across 49 human tissues using the gene name or SNP rs ID. It will return a table with the RefSeq

transcript ID, gene symbol, SNP rs ID, chromosome po- sition, tissue types, and *P*-value of each 3raQTL item for the queried gene or SNP. For example, querying IRF5 will return 12 308 significant 3raQTL items. Users can further

filter these 3raQTLs by selecting tissue names in the ‘Tis-

sues’ column (e.g. Whole Blood) or inputting custom fil-

ter key words (e.g. rs10954213) into the ‘Search’ field at the

top-right corner of the table (Figure [2](#_bookmark1)B). We also provide the ‘Browser’ and ‘Boxplot’ button for each 3raQTL item to allow users to visualize the 3raQTL item in the genome

browser and boxplot figures.

In the ‘3raQTL genome browser’ module, users can explore the 3raQTLs across human tissues in an in- teractive genome browser using the gene symbol (e.g.,

*IRF5*), SNP rs ID (e.g. rs10954213), or genome position (e.g. chr7:128687612–129200035). For example, by search-

ing ‘*SNCA*’ in the genome browser, we can find all sig- nificant 3raQTLs for *SNCA* (blue points) in Brain Cor- tex tissue, which confirms previous results that common

variants can change the APA usage of *SNCA* in the brain ([16](#_bookmark7)) (Figure [2](#_bookmark1)C). Clicking the dot of interest will show de- tails of the SNP, including rs ID and *P*-value (Figure [2](#_bookmark1)C).

Only 3raQTLs of the queried gene are labeled in blue in the genome browser, whereas 3raQTLs of other genes are labeled in grey. The genome browser also provides gene

structure annotation, GWAS catalog risk SNPs ([34](#_bookmark32)), and PolyA DB3 polyA sites ([38](#_bookmark28)) tracks, which allow users to

integrate these data with 3raQTLs. In addition, users can download the figures of the browser tracks in SVG format

by clicking on the ‘Save SVG’ button at the top-right corner of the genome browser.

In the ‘3raQTL boxplot’ module, we designed an on- line tool that allows users to customize boxplots for each 3raQTL. For example, users can draw the boxplot by in- putting the gene ID (e.g. NM 001347928@IRF5), rs ID

(e.g. rs10954213), and tissue name (e.g. Whole Blood) (Fig- ure [2](#_bookmark1)D). The color of the boxplot is user defined, and the whole plot can be downloaded as a publishable PDF docu- ment.

In the ‘GWAS-3raQTL colocalization event visualiza- tion’ module, we provide an online server for the widely

used R package LocusCompare ([37](#_bookmark27)), which allows users to visualize GWAS-3raQTLs colocalization events using their own GWAS data. For example, by inputting the gene

ID (e.g. NM 001382207@ZC3H13), tissue name (e.g. Pan-

creas), and a two-column text with the rs ID and corre- sponding *P*-value, users can visualize the GWAS-3raQTLs colocalization event at the *ZC3H13* locus (Figure [2](#_bookmark1)E). The

plots can also be downloaded in PDF format.

To link 3raQTL variants to human genetic traits and dis- eases, we also provide a list of GWAS-associated 3raQTLs, which were defined when the lead 3raQTL variants are the GWAS catalog ([34](#_bookmark32)) tag SNPs or SNPs in strong LD with

tag SNPs. This allows users to investigate the mechanisms of 3raQTLs in human traits and diseases.

# Downloading data and figures

We provide download functions for all four modules of the 3raQTL-atlas. In the ‘3raQTL search by Gene*/*SNP’ mod- ule, users can download a table file for all queried results. In the ‘3raQTL genome browser’ module, the genome browser can be downloaded in SVG format by clicking on the ‘Save

SVG’ button (e.g. Figure [2](#_bookmark1)C). For the other two modules, users can download a customized boxplot for each 3raQTL (e.g. Figure [2](#_bookmark1)D), as well as LocusCompare plots ([37](#_bookmark27)) for the GWAS-3raQTLs colocalization events, in PDF format [(e.g. Figure](https://wlcb.oit.uci.edu/3aQTLatlas/Download.php) [2](#_bookmark1)[E). We also provide a download page (https://](https://wlcb.oit.uci.edu/3aQTLatlas/Download.php)

[wlcb.oit.uci.edu/3aQTLatlas/Download.php), where users](https://wlcb.oit.uci.edu/3aQTLatlas/Download.php) can download all the 3raQTLs across 49 human normal tis- sues and a table of all human trait- and disease- associated 3raQTLs for further analysis.

# SUMMARY AND FUTURE DIRECTIONS

Increasing evidence suggests that genetic variants impact- ing APA usage play crucial roles in human diseases and traits ([22](#_bookmark16),[39](#_bookmark31),[40](#_bookmark33)). Here, we comprehensively evaluated the ef-

fects of genetic variants on 3rUTR usage across 49 human normal tissue types from the GTEx project and provide a user-friendly database, 3raQTL-atlas, for users to query, browse, visualize, and download the 3raQTLs. To the best of our knowledge, 3raQTL-atlas is the first database for users to explore the genetic effects on 3rUTR usage in large- scale human normal tissues. In recent years, similar QTL

resources, such as eQTL and sQTL, utilizing calculations from GTEx human normal tissues have been wildly used for

functional interpretations of GWAS risk loci. The 3raQTL- atlas aims to establish APA as another emerging and im-

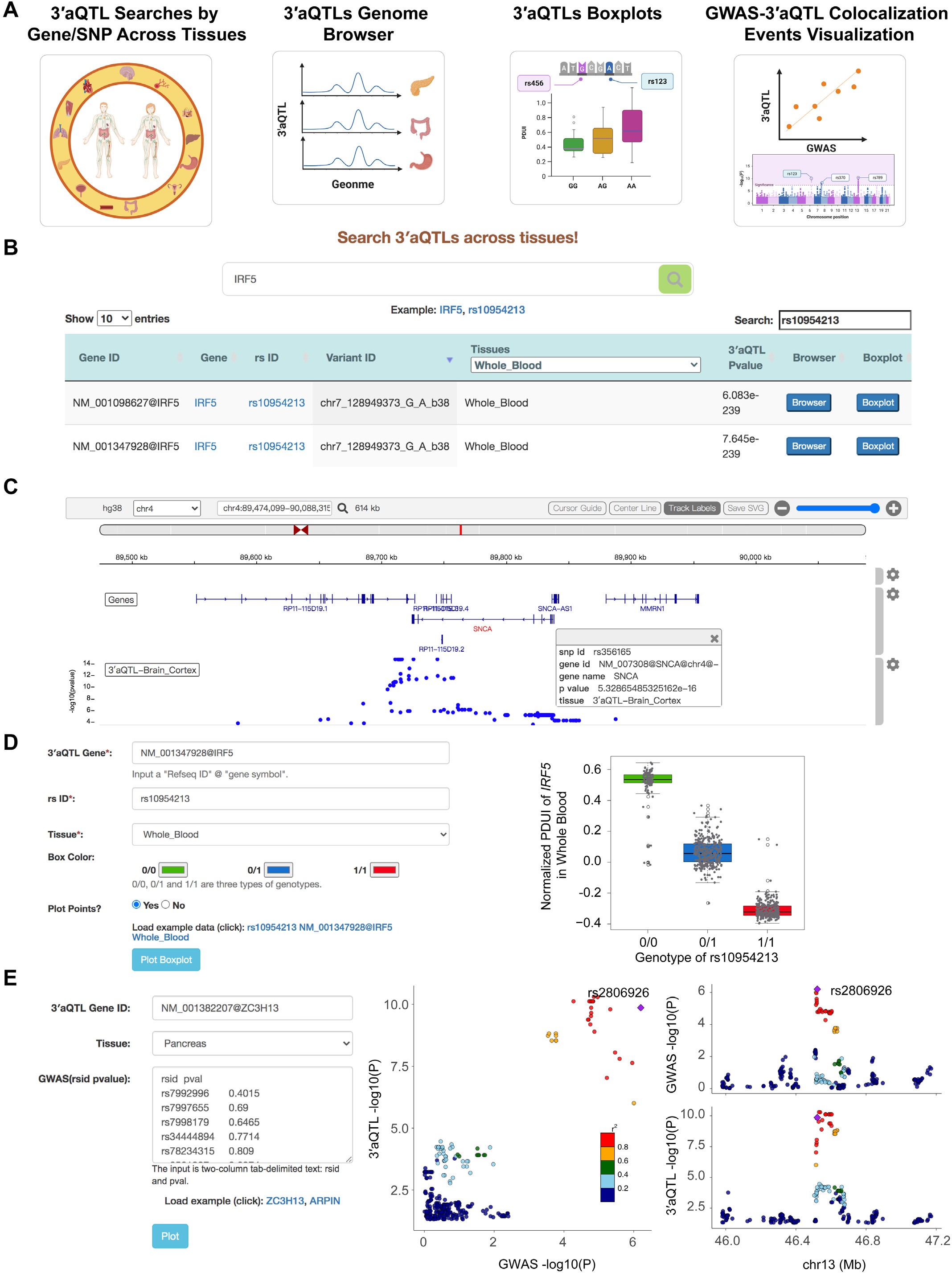
portant molecular phenotype to explain a large fraction of GWAS risk SNPs, leading to significant novel biological in- sights into the genetic basis of APA and APA-linked suscep- tibility genes in human traits and diseases. With the increas- ing number of RNA-seq datasets from the GTEx project and other consortium projects, such as the Trans-Omics for

Precision Medicine program ([41](#_bookmark35)), we will continue to up- date the 3raQTL-atlas to include 3raQTLs from more indi- viduals and tissue types. This would make the 3raQTL-atlas

an important resource for the genetic research community.

*Nucleic Acids Research, 2022, Vol. 50, Database issue* ***D43***

Downloaded from https://academic.oup.com/nar/article/50/D1/D39/6357732 by Baylor College of Medicine user on 01 September 2022



**Figure 2.** Web interface of 3raQTL-atlas. (**A**) 3raQTL-atlas consists of four modules: 3raQTL search by Gene*/*SNP, 3raQTL genome browser, 3raQTL box- plot, and GWAS-3raQTL colocalization event visualization. (**B**) 3raQTL query interface and sample results in the ‘3raQTL search by Gene*/*SNP’ module.

(**C**) An example of the genome browser view shows the 3raQTLs of brain cortex tissue at the SNCA locus. (**D**) Interface of the ‘3raQTL boxplot’ module and an example of the 3raQTL boxplot for *IRF5* and rs10954213 in whole blood. (**E**) Interface of the ‘GWAS-3raQTL colocalization event visualization’ module and an example of the LocusCompare plot at the *ZC3H13* locus with T2D GWAS *P*-values and 3raQTL *P*-values in pancreas tissue. GWAS, genome-wide association studies; SNP, single-nucleotide polymorphism; WGS, whole-genome sequencing; 3raQTL, 3r untranslated region alternative polyadenylation quantitative trait loci.

***D44*** *Nucleic Acids Research, 2022, Vol. 50, Database issue*

Downloaded from https://academic.oup.com/nar/article/50/D1/D39/6357732 by Baylor College of Medicine user on 01 September 2022

In summary, the 3raQTL-atlas provides significant supple- ments to interpret the function of non-coding GWAS risk

variants and offers beneficial resources for exploring the ge- netic basis of APA in human phenotypic diversity and a wide spectrum of human diseases.

# DATA AVAILABILITY

The data used for the analyses described in this manuscript were obtained from dbGaP accession number phs000424.v8.p2

# SUPPLEMENTARY DATA

[Supplementary Data](https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkab740#supplementary-data) are available at NAR Online.

# ACKNOWLEDGEMENTS

We thank members of the Li laboratory and Dr Jingyi Jes- sica Li for helpful discussions.

# FUNDING

The Genotype-Tissue Expression (GTEx) Project was sup- ported by the Common Fund of the Office of the Director of the National Institutes of Health; NCI, NHGRI, NHLBI, NIDA, NIMH and NINDS; National Institutes of Health [R01CA193466 to W.L.]. Funding for open access charge: National Institutes of Health.

*Conflict of interest statement.* None declared.

# REFERENCES

1. GTEx Consortium (2017) Genetic effects on gene expression across human tissues. *Nature*, **550**, 204–213.
2. GTEx Consortium (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, **369**, 1318–1330.
3. Chen,L., Ge,B., Casale,F.P., Vasquez,L., Kwan,T.,

Garrido-Martin,D., Watt,S., Yan,Y., Kundu,K., Ecker,S. *et al.* (2016) Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell*, **167**, 1398–1414.

1. Franzen,O., Ermel,R., Cohain,A., Akers,N.K., Di Narzo,A., Talukdar,H.A., Foroughi-Asl,H., Giambartolomei,C., Fullard,J.F., Sukhavasi,K. *et al.* (2016) Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science*, **353**, 827–830.
2. Gamazon,E.R., Segre,A.V., van de Bunt,M., Wen,X., Xi,H.S., Hormozdiari,F., Ongen,H., Konkashbaev,A., Derks,E.M., Aguet,F. *et al.* (2018) Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat.**Genet.*, **50**, 956–967.
3. Yao,D.W., O’Connor,L.J., Price,A.L. and Gusev,A. (2020) Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat. Genet.*, **52**, 626–633.
4. Hoque,M., Ji,Z., Zheng,D., Luo,W., Li,W., You,B., Park,J.Y., Yehia,G. and Tian,B. (2013) Analysis of alternative cleavage and

polyadenylation by 3r region extraction and deep sequencing. *Nat. Methods*, **10**, 133–139.

1. Elkon,R., Ugalde,A.P. and Agami,R. (2013) Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.*, **14**, 496–506.
2. Tian,B. and Manley,J.L. (2017) Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.*, **18**, 18–30.
3. Hong,W., Ruan,H., Zhang,Z., Ye,Y., Liu,Y., Li,S., Jing,Y., Zhang,H., Diao,L., Liang,H. *et al.* (2020) APAatlas: decoding alternative polyadenylation across human tissues. *Nucleic Acids Res.*, **48**, D34–D39.
4. Taliaferro,J.M., Vidaki,M., Oliveira,R., Olson,S., Zhan,L., Saxena,T., Wang,E.T., Graveley,B.R., Gertler,F.B., Swanson,M.S. *et al.* (2016) Distal alternative last exons localize mRNAs to neural projections. *Mol. Cell*, **61**, 821–833.
5. Jenal,M., Elkon,R., Loayza-Puch,F., van Haaften,G., Kuhn,U., Menzies,F.M., Oude Vrielink,J.A., Bos,A.J., Drost,J., Rooijers,K. *et al.* (2012) The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell*, **149**, 538–553.
6. Gennarino,V.A., Alcott,C.E., Chen,C.A., Chaudhury,A., Gillentine,M.A., Rosenfeld,J.A., Parikh,S., Wheless,J.W., Roeder,E.R., Horovitz,D.D. *et al.* (2015) NUDT21-spanning CNVs lead to neuropsychiatric disease and altered MeCP2 abundance via alternative polyadenylation. *Elife*, **4**, e10782.
7. Lee,S.H., Singh,I., Tisdale,S., Abdel-Wahab,O., Leslie,C.S. and Mayr,C. (2018) Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. *Nature*, **561**, 127–131.
8. Masamha,C.P., Xia,Z., Yang,J., Albrecht,T.R., Li,M., Shyu,A.B., Li,W. and Wagner,E.J. (2014) CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature*, **510**, 412–416.
9. Rhinn,H., Qiang,L., Yamashita,T., Rhee,D., Zolin,A., Vanti,W. and Abeliovich,A. (2012) Alternative alpha-synuclein transcript usage as a convergent mechanism in Parkinson’s disease pathology. *Nat. Commun.*, **3**, 1084.
10. Graham,R.R., Kyogoku,C., Sigurdsson,S., Vlasova,I.A., Davies,L.R., Baechler,E.C., Plenge,R.M., Koeuth,T., Ortmann,W.A., Hom,G. *et al.* (2007) Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 6758–6763.
11. Hellquist,A., Zucchelli,M., Kivinen,K., Saarialho-Kere,U., Koskenmies,S., Widen,E., Julkunen,H., Wong,A.,

Karjalainen-Lindsberg,M.L., Skoog,T. *et al.* (2007) The human GIMAP5 gene has a common polyadenylation polymorphism increasing risk to systemic lupus erythematosus. *J. Med. Genet.*, **44**, 314–321.

1. Stacey,S.N., Sulem,P., Jonasdottir,A., Masson,G., Gudmundsson,J., Gudbjartsson,D.F., Magnusson,O.T., Gudjonsson,S.A., Sigurgeirsson,B., Thorisdottir,K. *et al.* (2011) A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat. Genet.*, **43**, 1098–1103.
2. Garin,I., Edghill,E.L., Akerman,I., Rubio-Cabezas,O., Rica,I., Locke,J.M., Maestro,M.A., Alshaikh,A., Bundak,R., del Castillo,G. *et al.* (2010) Recessive mutations in the INS gene result in neonatal diabetes through reduced insulin biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 3105–3110.
3. Locke,J.M., Da Silva Xavier,G., Rutter,G.A. and Harries,L.W. (2011) An alternative polyadenylation signal in TCF7L2 generates isoforms that inhibit T cell factor*/*lymphoid-enhancer factor

(TCF*/*LEF)-dependent target genes. *Diabetologia*, **54**, 3078–3082.

1. Li,L., Huang,K.L., Gao,Y., Cui,Y., Wang,G., Elrod,N.D., Li,Y., Chen,Y.E., Ji,P., Peng,F. *et al.* (2021) An atlas of alternative polyadenylation quantitative trait loci contributing to complex trait and disease heritability. *Nat. Genet.*, **53**, 994–1005.
2. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
3. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
4. Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
5. Poplin,R., Ruano-Rubio,V., DePristo,M.A., Fennell,T.J., Carneiro,M.O., Van der Auwera,G.A., Kling,D.E., Gauthier,L.D., Levy-Moonshine,A., Roazen,D. *et al.* (2018) Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv doi: [https://doi.org/10.1101/201178](https://www.doi.org/10.1101/201178), 24 July 2018, preprint: not peer reviewed.
6. Delaneau,O., Zagury,J.F. and Marchini,J. (2013) Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods*, **10**, 5–6.
7. Feng,X., Li,L., Wagner,E.J. and Li,W. (2018) TC3A: the cancer 3r UTR atlas. *Nucleic Acids Res.*, **46**, D1027–D1030.
8. Karolchik,D., Hinrichs,A.S., Furey,T.S., Roskin,K.M., Sugnet,C.W., Haussler,D. and Kent,W.J. (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.*, **32**, D493–D496.

*Nucleic Acids Research, 2022, Vol. 50, Database issue* ***D45***

1. Lindenbaum,P. and Redon,R. (2018) bioalcidae, samjs and vcffilterjs: object-oriented formatters and filters for bioinformatics files. *Bioinformatics*, **34**, 1224–1225.

Downloaded from https://academic.oup.com/nar/article/50/D1/D39/6357732 by Baylor College of Medicine user on 01 September 2022

1. Shabalin,A.A. (2012) Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*, **28**, 1353–1358.
2. Stegle,O., Parts,L., Piipari,M., Winn,J. and Durbin,R. (2012) Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat.* *Protoc.*, **7**, 500–507.
3. Storey,J.D. and Tibshirani,R. (2003) Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 9440–9445.
4. MacArthur,J., Bowler,E., Cerezo,M., Gil,L., Hall,P., Hastings,E., Junkins,H., McMahon,A., Milano,A., Morales,J. *et al.* (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.*, **45**, D896–D901.
5. McClellan,J. and King,M.C. (2010) Genetic heterogeneity in human disease. *Cell*, **141**, 210–217.
6. Cowper-Sal lari,R., Zhang,X., Wright,J.B., Bailey,S.D., Cole,M.D.,

Eeckhoute,J., Moore,J.H. and Lupien,M. (2012) Breast cancer

risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat. Genet.*, **44**, 1191–1198.

1. Liu,B., Gloudemans,M.J., Rao,A.S., Ingelsson,E. and Montgomery,S.B. (2019) Abundant associations with gene expression complicate GWAS follow-up. *Nat. Genet.*, **51**, 768–769.
2. Wang,R., Nambiar,R., Zheng,D. and Tian,B. (2018) PolyA DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. *Nucleic Acids Res.*, **46**, D315–D319.
3. Zheng,Z., Huang,D., Wang,J., Zhao,K., Zhou,Y., Guo,Z., Zhai,S., Xu,H., Cui,H., Yao,H. *et al.* (2020) QTLbase: an integrative resource for quantitative trait loci across multiple human molecular phenotypes. *Nucleic Acids Res.*, **48**, D983–D991.
4. Yang,Y., Zhang,Q., Miao,Y.R., Yang,J., Yang,W., Yu,F., Wang,D., Guo,A.Y. and Gong,J. (2020) SNP2APA: a database for evaluating effects of genetic variants on alternative polyadenylation in human cancers. *Nucleic Acids Res.*, **48**, D226–D232.
5. Taliun,D., Harris,D.N., Kessler,M.D., Carlson,J., Szpiech,Z.A., Torres,R., Taliun,S.A.G., Corvelo,A., Gogarten,S.M., Kang,H.M. *et al.* (2021) Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*, **590**, 290–299.