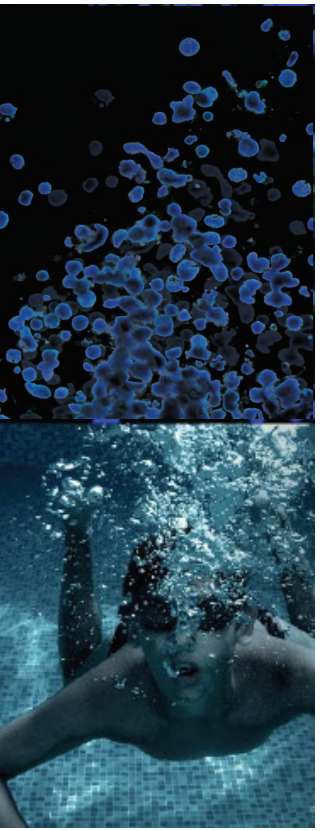
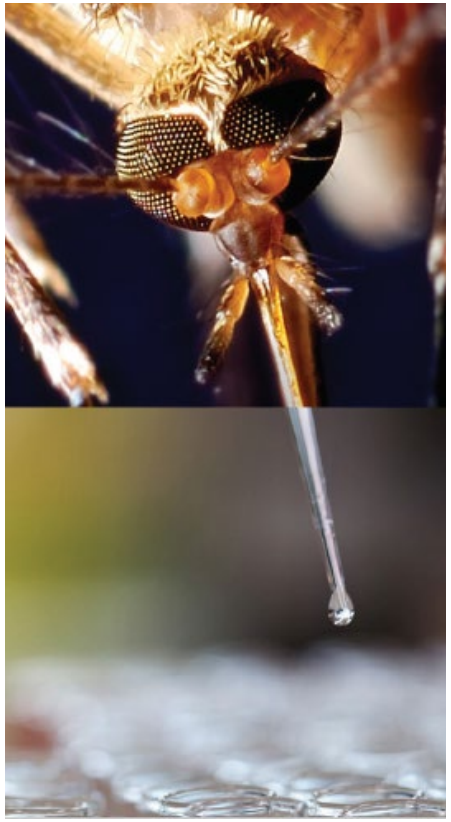
Characterizing the Molecular Qualities of Authenticated ATCC® Cell Lines

Implementing scientific rigor, reproducibility, and innovation in the preclinical investigations

Ajeet P. Singh, PhD
Senior Scientist, Bioinformatics
Sequencing & Bioinformatics Center
ATCC

Credible Leads to Incredible™





About ATCC®

- Founded in 1925
- 501(c)(3) not-for-profit organization
- World's largest, most diverse biorepository
- Quality accreditation by multiple industry standards
  - ISO 9001 Certified
  - ISO 13485 Certified
  - ISO/IEC 17025 Accredited
  - ISO 17034 Accredited

- Standards development partner with multiple industry working groups
  - ANSI Standards Working Groups
  - AOAC International Working Group
  - IMMSA/NIST Microbiome Standards
- Global supplier of authenticated cell lines, microorganisms, and molecular standards
- Sales and distribution to 150+ countries
- Talented team of 500+ employees

Genomics data quality

Connecting the dots between bioinformatics and physical materials

- Review challenges associated with genomics data quality and authenticity
- Discuss ***why*** ATCC® is committed to **providing reference-quality transcriptomes** for our cell lines
- Discuss our current efforts to produce standardized transcriptomes reference data
- Explore the ATCC® Cell Line Land

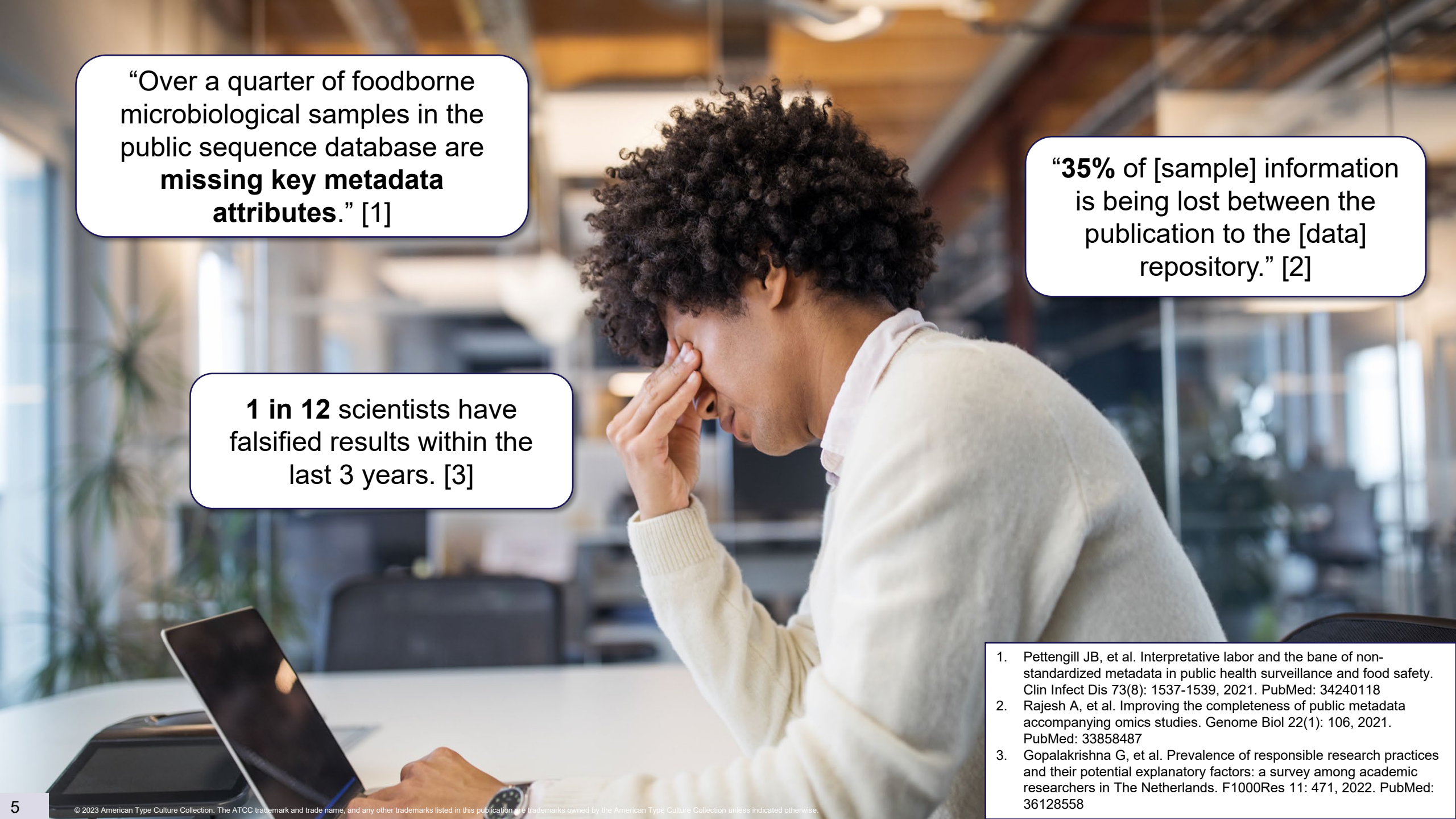
Challenges stemming from poor data quality…

"***Finding the right cell lines*** for my research is a challenge."

"Many cell types are ***not good models*** for the disease I'm studying."

"Pre-existing results are difficult to reproduce and often ***not reproducible***."

ATCC



"Over a quarter of foodborne microbiological samples in the public sequence database are **missing key metadata attributes**." [1]

"**35%** of [sample] information is being lost between the publication to the [data] repository." [2]

1 in 12 scientists have falsified results within the last 3 years. [3]

1. Pettengill JB, et al. Interpretative labor and the bane of non-standardized metadata in public health surveillance and food safety. Clin Infect Dis 73(8): 1537-1539, 2021. PubMed: 34240118
2. Rajesh A, et al. Improving the completeness of public metadata accompanying omics studies. Genome Biol 22(1): 106, 2021. PubMed: 33858487
3. Gopalakrishna G, et al. Prevalence of responsible research practices and their potential explanatory factors: a survey among academic researchers in The Netherlands. F1000Res 11: 471, 2022. PubMed: 36128558

Fake data was first discovered in GenBank in 1997

"Mr. Hajra, former graduate student, University of Michigan, engaged in scientific misconduct by falsifying and fabricating research data in five published research papers, two published review articles, one submitted but unpublished paper, in his doctoral dissertation, and in a ***submission to the GenBank computer data base.****" – The Federal Register, v62, n135 (****1997****)*

Federal Register / Vol. 62, No. 135 / Tuesday, July 15, 1997 / Notices 37921

author of the application is identified
n's role in the project is
nts
l Experience. The
es the qualifying
ganization to
licant's ability to
iently administer
lication specifically
ant as a nationally-
tion, institution, or
ord of study and
l special
s. Previous specific
rk similar to the
learly and
ed. The relationship
t and other work
d, or underway by
cribed, including a
l related Federal
ed within the last five
years. In the event a consortium of applicants is proposed, the project history of prior joint work should be provided. The previous Federal assistance is identified by project number, Federal agency, and grants or contracting officer. 25 points

Components of a Complete Application

A complete application consists of the following items in this order:

1. Application for Federal Assistance (Standard Form 424, REV 4–88);
2. Budget Information—Non-construction Programs (Standard Form 424A, REV 4–88);
3. Assurances—Non-construction Programs (Standard Form 424B, REV 4–88);
4. Table of Contents;

Dated: July 9, 1997.

David F. Garrison,

Principal Deputy Assistant Secretary for Planning and Evaluation.

[FR Doc. 97–18528 Filed 7–14–97; 8:45 am]

BILLING CODE 4151–04–M

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Office of the Secretary

Findings of Scientific Misconduct

AGENCY: Office of the Secretary, HHS.

ACTION: Notice.

SUMMARY: Notice is hereby given that the Office of Research Integrity (ORI) has made a final finding of scientific misconduct in the following case:

Amitav Hajra, University of Michigan: Based upon a report from the University of Michigan, information obtained by the Office of Research Integrity (ORI) during its oversight review, and Mr. Hajra's own admission, ORI found that Mr. Hajra, former graduate student, University of Michigan, engaged in scientific misconduct by falsifying and fabricating research data in five published research papers, two published review articles, one submitted but unpublished paper, in his doctoral dissertation, and in a submission to the GenBank computer data base. Mr. Hajra's doctoral training and research was supported by two Public Health Service (PHS) grants, and his experiments were conducted at and submitted for publication from the

- Wijmenga, C., Gregory, P.E., Hajra, A., Schröck, E., Ried, T., Eils, R., Liu, P.P., and Collins, F.S. "Core binding factor β-smooth muscle myosin heavy chain chimeric protein involved in acute myeloid leukemia forms unusual nuclear rod-like structures in transformed NIH 3T3 cells." *Proc. Natl. Acad. Sci.* USA 93(4):1630–1635, 1996; and
- Liu, P.P., Wijmenga, C., Hajra, A., Blake, T.B., Kelley, C.A., Adelstein, R.S., Bagg, A., Rector, J., Cotelingham, J., Willman, C.L., and Collins, F.S. "Identification of the chimeric protein product of the CBFB-MYH11 fusion gene in inv(16) leukemia cells." *Genes, Chromosomes, and Cancer* 16:77–87, 1996 (Erratum in *Genes, Chromosomes, and Cancer* 18(1):71, 1997).

Mr. Hajra included fabricated and falsified data in the following review articles:

- Hajra, A., Liu, P.P., and Collins, F.S. "Transforming properties of the leukemic Inv(16) fusion gene CBFB–MYH11." In Molecular Aspects of Myeloid Stem Cell Development in Current Topics in Microbiology and Immunology (L. Wolff and A.S. Perkins, Eds.) 211:289–298, 1996 (Review). Berlin and New York: Springer-Verlag; and
- Liu, P.P., Hajra, A., Wijmenga, C., and Collins, F.S. "Molecular pathogenesis of the chromosome 16 inversion in the M4Eo subtype of acute myeloid leukemia." *Blood* 85:2289–2302, 1995 (Review).

Mr. Hajra submitted a fabricated nucleotide sequence in computer data

24 years later, this falsified data is still being cited…

Received: 25 March 2021 | Revised: 16 June 2021 | Accepted: 13 July 2021
DOI: 10.1002/humu.24261

REVIEW

Human Mutation

Pathogenic noncoding variants in the neurofibromat schwannomatosis predisposition genes

Cristina Perez-Becerril

Division of Evolution and Genomic Science, Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester Academic Health Science Centre, School of Biological Sciences, University of Manchester, Manchester, UK

Correspondence
Miriam J. Smith, Division of Evolution and Genomic Science, Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester Academic Health Science Centre, School of Biological Sciences, University of Manchester, Manchester M13 9WL, UK.
Email: miriam.smith@manchester.ac.uk

comparison of the full human and murine neurofibromin sequences revealed a high degree of similarity (>98%) and high conservation levels across 5′- and 3′-UTRs (Bernards et al., 1993; Hajra et al., 1994). A subsequent *in silico* study compared the 5′ upstream region and intron 1 of *NF1* and homologous genes in human, mouse, rat, and puffer fish (*Fugu rubripes*). The authors found high homology segments throughout the region across all species, including two exact

tumors. Pathogenic variants (PVs) that cause NF1 and NF2 are

and *NF2* loci, respectively. To date, most variants associated wi

have been identified in the *SMARCB1* and *LZTR1* genes, and a

the *DGCR8* gene was recently reported to predispose to schwa

the high detection rate for PVs in NF1 and NF2 (over 90% of r

variants can be identified by routine genetic screening) under

portion of clinical cases remain undetected. A higher prop

Federal Register / Vol. 62, No. 135 / Tuesday, July 15, 1997 / Notices 37921

author of the application is identified and that person's role in the project is identified. 20 points

4. *Organizational Experience.* The application identifies the qualifying experience of the organization to demonstrate the applicant's ability to effectively and efficiently administer this project. The application specifically identifies the applicant as a nationally-recognized organization, institution, or company with a record of study and analysis of rural and special transportation needs. Previous specific experience with work similar to the Tasks proposed is clearly and specifically described. The relationship between this project and other work planned, anticipated, or underway by the applicant is described, including a chart which lists all related Federal assistance received within the last five years. In the event a consortium of applicants is proposed, the project history of prior joint work should be provided. The previous Federal assistance is identified by project number, Federal agency, and grants or contracting officer. 25 points

Components of a Complete Application

A complete application consists of the following items in this order:

1. Application for Federal Assistance (Standard Form 424, REV 4–88);
2. Budget Information—Non-construction Programs (Standard Form 424A, REV 4–88);
3. Assurances—Non-construction Programs (Standard Form 424B, REV 4–88);
4. Table of Contents;

Dated: July 9, 1997.

David F. Garrison,

Principal Deputy Assistant Secretary for Planning and Evaluation.

[FR Doc. 97–18528 Filed 7–14–97; 8:45 am]

BILLING CODE 4151–04–M

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Office of the Secretary

Findings of Scientific Misconduct

AGENCY: Office of the Secretary, HHS.

ACTION: Notice.

SUMMARY: Notice is hereby given that the Office of Research Integrity (ORI) has made a final finding of scientific misconduct in the following case:

Amitav Hajra, University of Michigan: Based upon a report from the University of Michigan, information obtained by the Office of Research Integrity (ORI) during its oversight review, and Mr. Hajra's own admission, ORI found that Mr. Hajra, former graduate student, University of Michigan, engaged in scientific misconduct by falsifying and fabricating research data in five published research papers, two published review articles, one submitted but unpublished paper, in his doctoral dissertation, and in a submission to the GenBank computer data base. Mr. Hajra's doctoral training and research was supported by two Public Health Service (PHS) grants, and his experiments were conducted at and submitted for publication from the

- Wijmenga, C., Gregory, P.E., Hajra, A., Schröck, E., Ried, T., Eils, R., Liu, P.P., and Collins, F.S. "Core binding factor β-smooth muscle myosin heavy chain chimeric protein involved in acute myeloid leukemia forms unusual nuclear rod-like structures in transformed NIH 3T3 cells." *Proc. Natl. Acad. Sci.* USA 93(4):1630–1635, 1996; and
- Liu, P.P., Wijmenga, C., Hajra, A., Blake, T.B., Kelley, C.A., Adelstein, R.S., Bagg, A., Rector, J., Cotelingham, J., Willman, C.L., and Collins, F.S. "Identification of the chimeric protein product of the CBFB-MYH11 fusion gene in inv(16) leukemia cells." *Genes, Chromosomes, and Cancer* 16:77–87, 1996 (Erratum in *Genes, Chromosomes, and Cancer* 18(1):71, 1997).

Mr. Hajra included fabricated and falsified data in the following review articles:

- Hajra, A., Liu, P.P., and Collins, F.S. "Transforming properties of the leukemic Inv(16) fusion gene CBFB–MYH11." In Molecular Aspects of Myeloid Stem Cell Development in Current Topics in Microbiology and Immunology (L. Wolff and A.S. Perkins, Eds.) 211:289–298, 1996 (Review). Berlin and New York: Springer-Verlag; and
- Liu, P.P., Hajra, A., Wijmenga, C., and Collins, F.S. "Molecular pathogenesis of the chromosome 16 inversion in the M4Eo subtype of acute myeloid leukemia." *Blood* 85:2289–2302, 1995 (Review).

Mr. Hajra submitted a fabricated nucleotide sequence in computer data

ATCC

After 42 citations… the data is still in GenBank…

DNA Sequences in the Promoter

https://www.sciencedirect.com/science/article/abs/pii/S08887543...

View PDF | Access through your institution | Purchase PDF | Search ScienceDirect

Article preview

Abstract

Cited by (42)

Recommended articles (6)

Short Communication

DNA Sequences in the Promoter Region of the *NF1* Gene Are Highly Conserved between Human and Mouse

Amitav Hajra, Antonia Martin-Gallardo, Susan A. Tarlé, Matthew Freedman, Susan Wilson-Gunn, Andre Bernards, Francis S. Collins

Show more

+ Add to Mendeley | Share | Cite

https://doi.org/10.1006/geno.1994.1328 Get rights and content

Abstract

The gene for type 1 neurofibromatosis (*NF1*) is most highly expressed in brain and spinal cord, although low levels of mRNA can be found in nearly all tissues. As a first step in investigating the regulation of *NF1* gene expression, we have cloned and sequenced the promoter regions of the human and mouse *NF1* genes and mapped the transcriptional start sites in both species. We report here that the 5′ ends of the human and murine *NF1* genes are highly conserved. While no discernable TATA or CCAAT box sequences are seen, transcription initiates at identical sites in both species, 484 nucleotides upstream of the ATG initiation codon in the human gene. The human and mouse *NF1* genes share particularly high sequence homology (95%) between nucleotides -33 and +261 and contain several perfectly conserved transcription factor binding site motifs, including a cAMP response element, several AP2 consensus binding sites, and a serum response element. The high conservation of these sequences indicates that they are likely to be significant in the regulation of *NF1* gene expression.

References (0)

Cited by (42)

Synthetic promoter for efficient and muscle-specific expression of exogenous genes
2019, Plasmid

Citation Excerpt :

…To acquire synthetic promoters with activity and specificity high enough for therapeutic application, we speculated that the combination of multiple kinds of transcriptional motifs, instead of unique myogenic

FEED

Human neurofibromin (NF1) ge

https://www.ncbi.nlm.nih.gov/nuccore/U17084

GenBank | Send to:

Human neurofibromin (NF1) gene, promoter region and partial cds

GenBank: U17084.1

FASTA | Graphics

Go to:

```
LOCUS       HSU17084                3953 bp    DNA     linear   PRI 07-DEC-1994
DEFINITION  Human neurofibromin (NF1) gene, promoter region and partial cds.
ACCESSION   U17084 U09106
VERSION     U17084.1
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 2943 to 3953)
  AUTHORS   Hajra,A., Martin-Gallardo,A., Tarle,S.A., Freedman,M.,
            Wilson-Gunn,S., Bernards,A. and Collins,F.S.
  TITLE     DNA sequences in the promoter region of the NF1 gene are highly
            conserved between human and mouse
  JOURNAL   Genomics 21 (3), 649-652 (1994)
   PUBMED   7959746
REFERENCE   2  (bases 1 to 3953)
  AUTHORS   Marchuk,D.A., Saulino,A.M., Tavakkol,R., Swaroop,M., Wallace,M.R.,
            Andersen,L.B., Mitchell,A.L., Gutmann,D.H., Boguski,M. and
            Collins,F.S.
  TITLE     cDNA cloning of the type 1 neurofibromatosis gene: complete
            sequence of the NF1 gene product
  JOURNAL   Genomics 11 (4), 931-940 (1991)
   PUBMED   1783401
REFERENCE   3  (bases 1 to 3953)
  AUTHORS   Hajra,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-NOV-1994) Amitav Hajra, Laboratory of Gene Transfer,
            National Center for Human Genome Research, NIH, Building 49, Room
            3A23, 49 Convent Drive, MSC 4470, Building 49, Room 3A23, 9000
```

Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

More about the gene LOC111811965

This region represents the pr
associated regulatory elemer
neurofibromin 1 gene. It inclu
sequence downstream of the

Related information

Protein

PubMed

Taxonomy

Gene

OMIM

GEO Profiles

LinkOut to external resou

ATCC

Data irreproducibility – a big toll on economics and credibility

PLOS | BIOLOGY

Prevalence of irreproducible preclinical research was greater than 50%, correlating with the expenditure of $28 billion per year in the United States on basic biomedical research that cannot be repeated successfully.

PERSPECTIVE

The Economics of Reproducibility in Preclinical Research

Leonard P. Freedman[1]*, Iain M. Cockburn[2], Timothy S. Simcoe[2,3]

1 Global Biological Standards Institute, Washington, D.C., United States of America, 2 Boston University School of Management, Boston, Massachusetts, United States of America, 3 Council of Economic Advisers, Washington, D.C., United States of America

* lfreedman@gbsi

The use of poor biological reagents and reference materials contributed the most to the reproducibility problem at 36.1%.

Abstract

Low reproducibi
production and contribute to both delays and costs of therapeutic drug development. An analysis of past studies indicates that the cumulative (total) prevalence of irreproducible preclinical research exceeds 50%, resulting in approximately US$28,000,000,000 (US $28B)/year spent on preclinical research that is not reproducible—in the United States alone. We outline a framework for solutions and a plan for long-term improvements in reproducibility rates that will help to accelerate the discovery of life-saving therapies and cures.

CrossMark
click for updates

OPEN ACCESS

Citation: Freedman LP, Cockburn IM, Simcoe TS (2015) The Economics of Reproducibility in Preclinical Research. PLoS Biol 13(6): e1002165. doi:10.1371/journal.pbio.1002165

ATCC

Challenging traceability of most public genomics data

1970 | 1980 | 1990 | 2000 | 2010 | 2020

Benton

??Transfer

??Transfer

??Transfer

Combs

3 decades of unknown transfers & research

NCBI

2009: Initial Deposit
"Draft" Reference Genome
ATCC XYZ1

1974: Strain Deposit
ATCC XYZ1

ATCC®

1995 MTA

Yarmosh

2 decades of lab research

NCBI

2015: Updated Genome
"High-Quality" Assembly
ATCC XYZ1

Authenticated source material

ATCC® Genome Portal Assembly

ATCC®

2020: Authenticated Genome Reference
ATCC XYZ1

100% authenticity and traceability

Names changed to protect the innocent

Potential issues with genomic source material

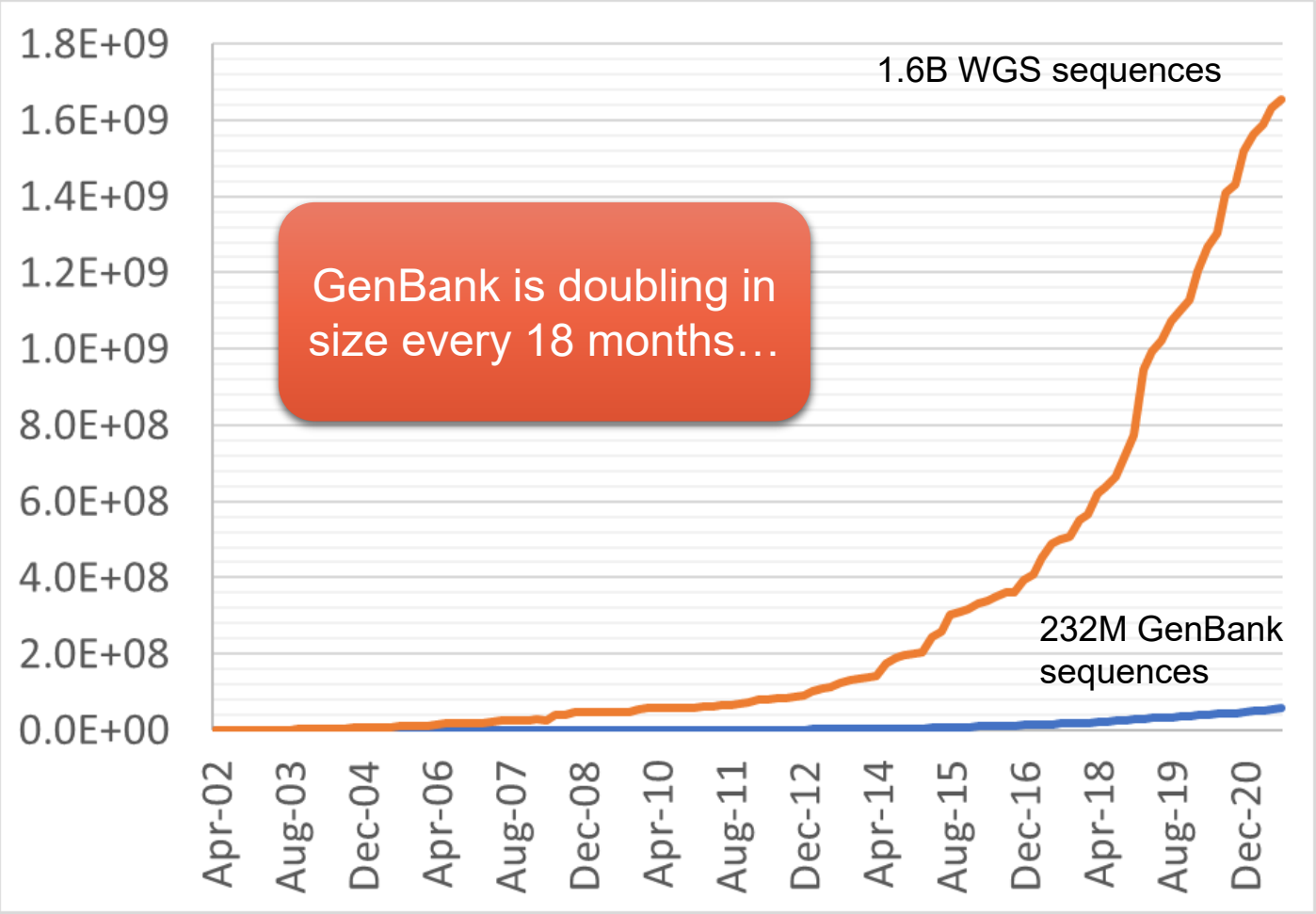
- "Lab adaptation"
- Loss of plasmids
- Sample mix ups
- Unknown chain of custody
- Differences in sequencing technology and bioinformatics

SNPs differences	ATCC Reference
Draft genome	>13,000
"High-quality" reference genome	>10,000

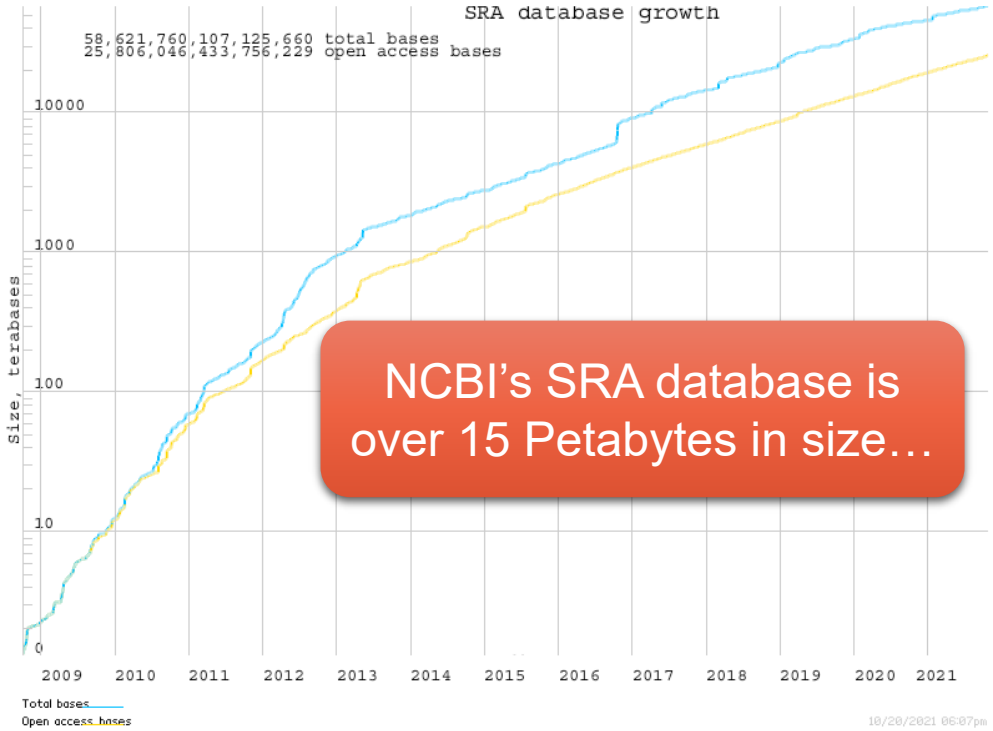
ATCC®

A reminder on the growth of public genomics data

1.6B sequences in WGS
232M sequences in GenBank



GenBank is doubling in size every 18 months…



NCBI's SRA database is over 15 Petabytes in size…

Data curation is a huge challenge

ATCC

Genomics data quality issues impact many disciplines

Factors

- Misclassification of sequences
- Chimeric genome assemblies
- Sample contamination
- Sequencing errors
- Mislabeling or data errors
- Data omission
- Data obfuscation
- Intentional misconduct

Critically Impacted Areas

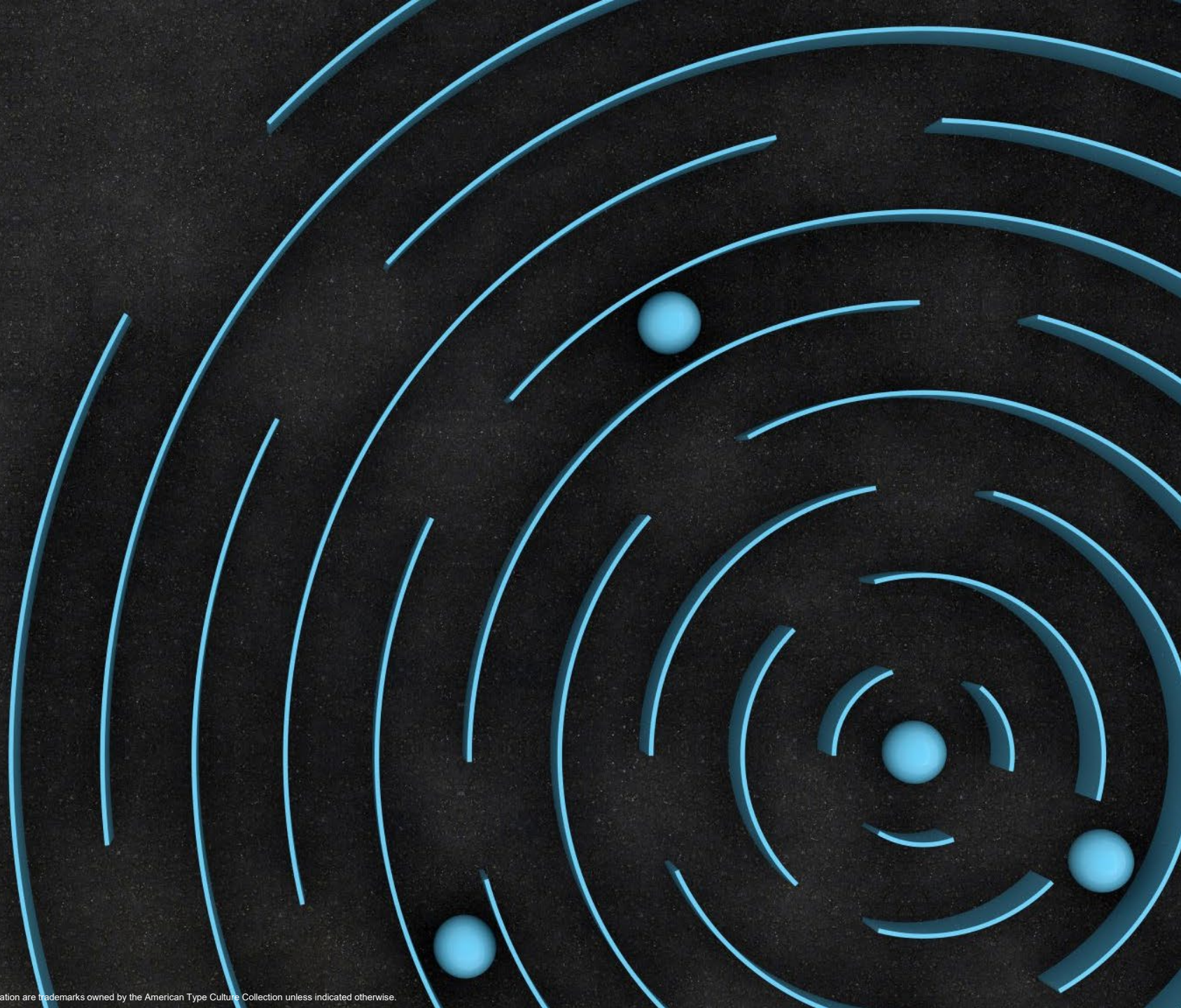
- Basic research (hypothesis generation)
- Biodiversity and environmental sciences
- Diagnostics & epidemiology
- Forensics
- Food safety
- Biodefense
- Many other areas…

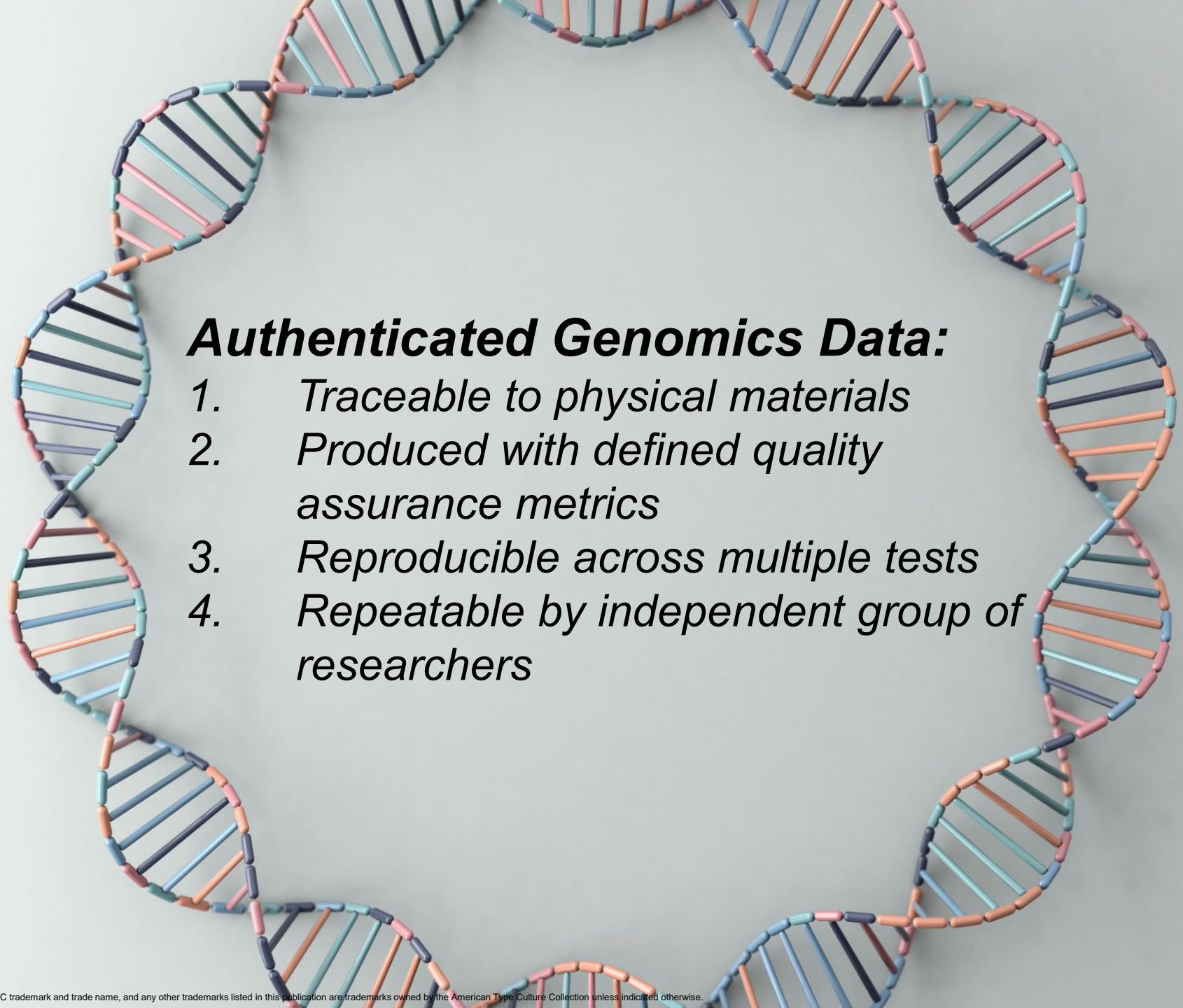
These are not “new” problems.

Many groups have sought solutions.

None, however, have sought to create **Authenticated Genomics Data**

What is “Authenticated Genomics Data”?





Authenticated Genomics Data:

1. *Traceable to physical materials*
2. *Produced with defined quality assurance metrics*
3. *Reproducible across multiple tests*
4. *Repeatable by independent group of researchers*

Journey of the data curation and their limitations

Lack traceability of the actual source material and data provenance

GenBank
ENA
SRA
EGA
GEO

Uncontrolled Public Data

- Unknown quality
- Missing or non-standard metadata
- Risky to use

Focused Public Data

- Improved metadata
- Moderate risk
- Often access-controlled
- Limited scope

Expert Curated Data

- Standardized metadata
- Standardized biofx methods
- Improved reproducibility
- Less risk, more results
- FAIR data model

Authenticated Data

- Standardized laboratory methods
- Quality Assurance (ISO)
- Traceable to materials in a biorepository
- ***Maximum data provenance***
- ***Maximum reproducibility***

ATCC

Journey of the data curation and their limitations

Often reproducible, gaps in data provenance

RefSeq
proGenomes
dbGAP
BluePrint
ICGC
TCGA
TARGET
CCLE
GTEx

Uncontrolled Public Data

- Unknown quality
- Missing or non-standard metadata
- Risky to use

Focused Public Data

- Improved metadata
- Moderate risk
- Often access-controlled
- Limited scope

Expert Curated Data

- Standardized metadata
- Standardized biofx methods
- Improved reproducibility
- Less risk, more results
- FAIR data model

Authenticated Data

- Standardized laboratory methods
- Quality Assurance (ISO)
- Traceable to materials in a biorepository
- ***Maximum data provenance***
- ***Maximum reproducibility***

ATCC

Journey of the data curation and their limitations

Gaps in data provenance

OmicSoft
Ingenuity (IPA)
HGMD

QIAGEN

Authenticated Data

- Standardized laboratory methods
- Quality Assurance (ISO)
- Traceable to materials in a biorepository
- ***Maximum data provenance***
- ***Maximum reproducibility***

Expert Curated Data

- Standardized metadata
- Standardized biofx methods
- Improved reproducibility
- Less risk, more results
- FAIR data model

Focused Public Data

- Improved metadata
- Moderate risk
- Often access-controlled
- Limited scope

Uncontrolled Public Data

- Unknown quality
- Missing or non-standard metadata
- Risky to use

ATCC

Authenticated transcriptomics data at ATCC

ATCC® is focused on data provenance and closing the reproducibility gap

OmicSoft
Ingenuity (IPA)
HGMD

QIAGEN

ATCC® Genome Portal
ATCC® Cell Line Land

ATCC®

Authenticated Data

- Standardized laboratory methods
- Quality Assurance (ISO)
- Traceable to materials in a biorepository
- ***Maximum data provenance***
- ***Maximum reproducibility***

Expert Curated Data

- Standardized metadata
- Standardized biofx methods
- Improved reproducibility
- Less risk, more results
- FAIR data model

Focused Public Data

- Improved metadata
- Moderate risk
- Often access-controlled
- Limited scope

Uncontrolled Public Data

- Unknown quality
- Missing or non-standard metadata
- Risky to use

ATCC®

The ATCC® Cell Line Land

A partnership with QIAGEN® Digital Insights: Enhancing scientific rigor and tackling the reproducibility gap

Why authenticated biomaterial is needed

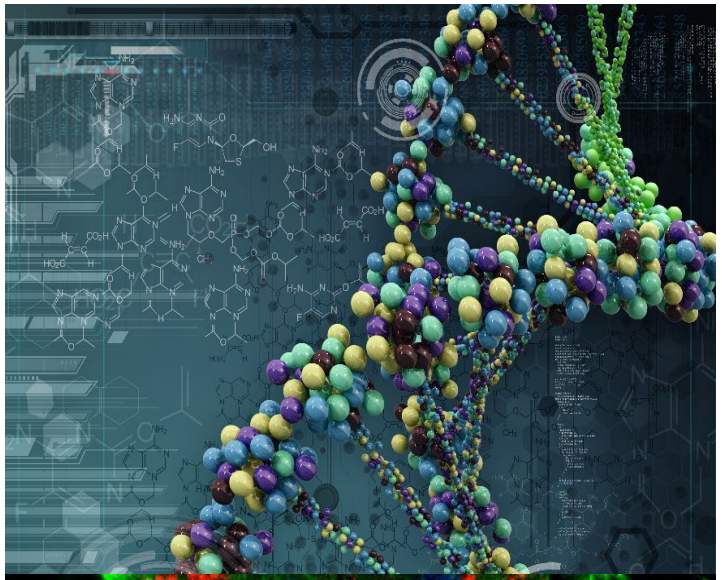
ATCC®, a trusted source of your authenticated cell lines

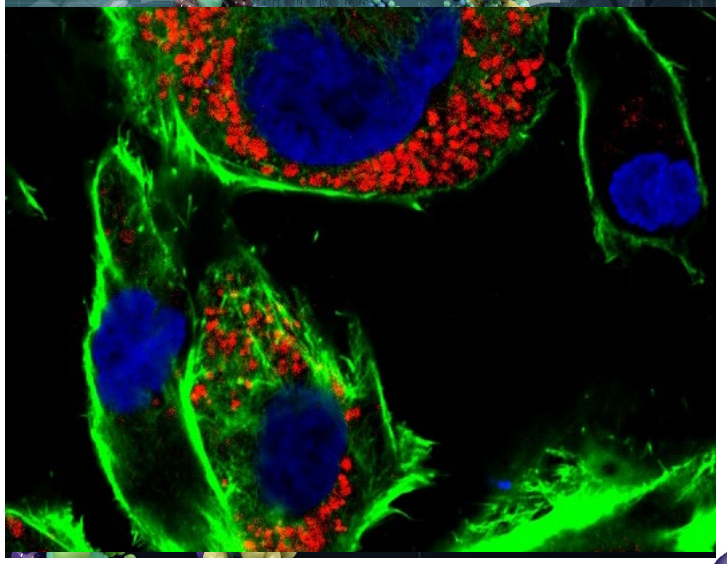
- An estimated 15-20% of all experiments found in the literature are using misidentified cell lines.
- Short tandem repeat (STR) profiling represents the gold standard for cell line authentication; however, cell lines (eg, HEK293) with acquired mutations in mismatch repair genes can alter their characteristic repeats upon prolonged culture, leading to negative authentication results via STR.
- In order to combat negative authentication results, ATCC® is completing RNA-seq analysis on their broad selection of human kidney cells using well-established protocols and stringent quality metrics.
- All fully authenticated human and mouse cell lines whole transcriptomics data will be available through ATCC® Cell Line Land.
- For more information: https://digitalinsights.qiagen.com/atcc-cell-line-land.

ATCC

ATCC® cell biology collection

ATCC® has **3,000+ authenticated** mammalian cell lines, genetic engineered cell lines, primary cells, stem cells, iPSCs, hTERT-immortalized cells, and organoids representing various species, cell types, tissues origins, and diseases.



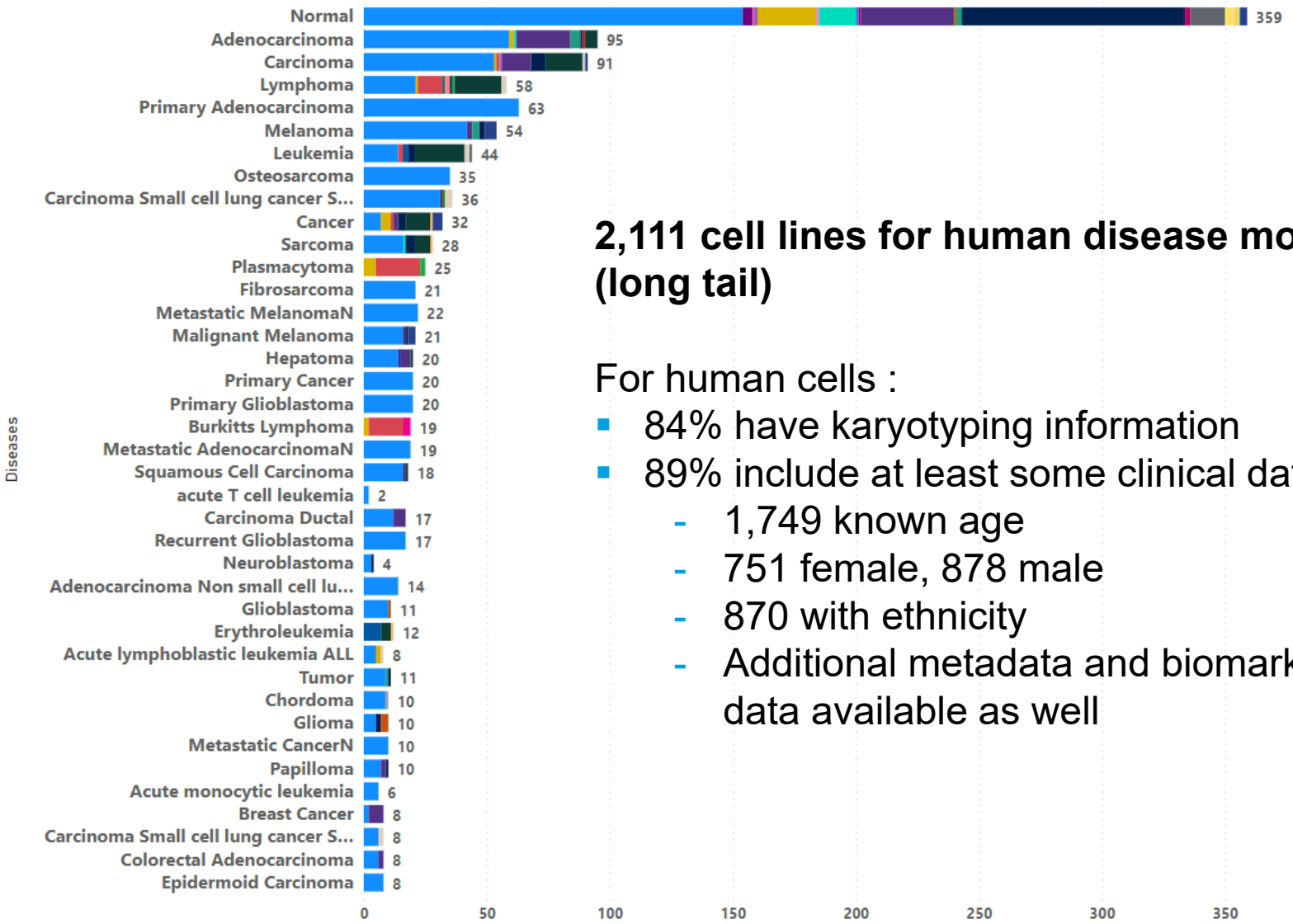


70+	100+	100+	400+
Species	Cell types	Tissue types	Diseases types

ATCC

ATCC® cell biology collection (by disease type)

Cell line models for over 400 disease types



2,111 cell lines for human disease models (long tail)

For human cells :

- 84% have karyotyping information
- 89% include at least some clinical data
  - 1,749 known age
  - 751 female, 878 male
  - 870 with ethnicity
  - Additional metadata and biomarker data available as well

ATCC

ATCC® Cell Line Land

A partnership with QIAGEN® Digital Insights

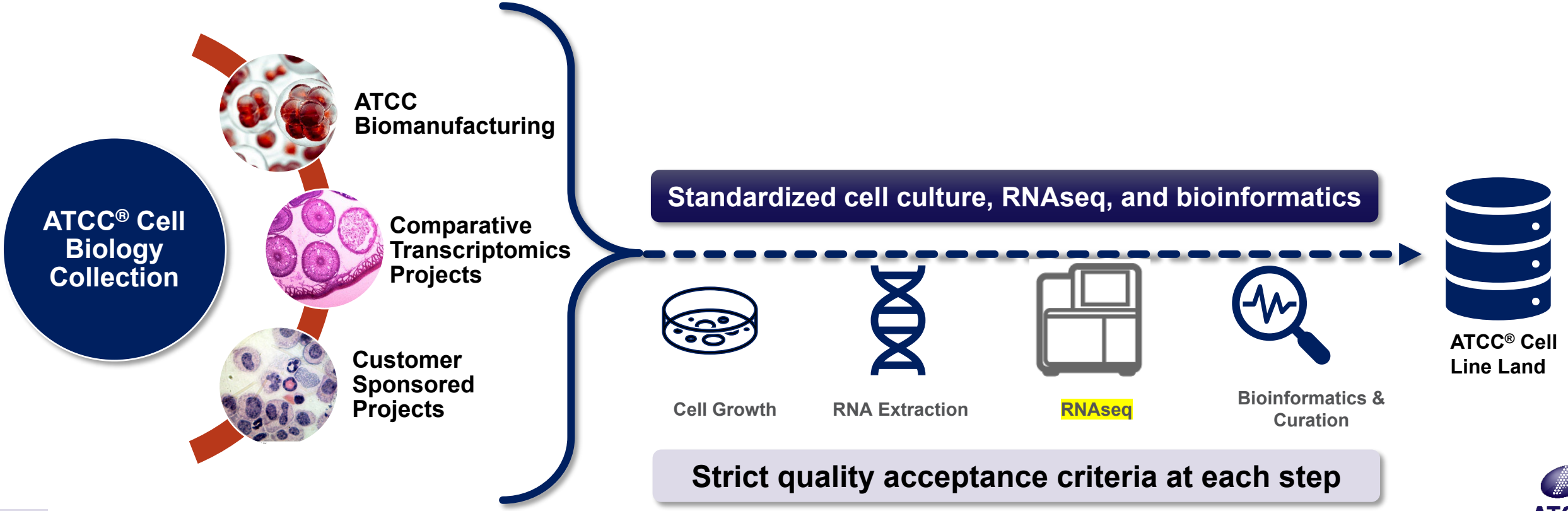
Q4 2022	Q1 2023	Q2 2023	Q3 2023	Q4 2023
Kidney	Blood Lymph Spleen	Lung Liver	Liver Brain	TBD
+250 reference data sets	+250 reference data sets	+250 reference data sets	+250 reference data sets	+250 reference data sets

- Current road-map for data production is subject to change
- Based on customer feedback
- **1,000+ traceable, authenticated RNAseq datasets per year**

ATCC

ATCC® Cell Line Land

Key Features	1. Repository of ***authenticated 'omics data traceable to physical materials*** 2. Data production, curation, and analysis uniformly standardized 3. Enables the highest level of **scientific reproducibility** 4. End-to-end **data provenance**



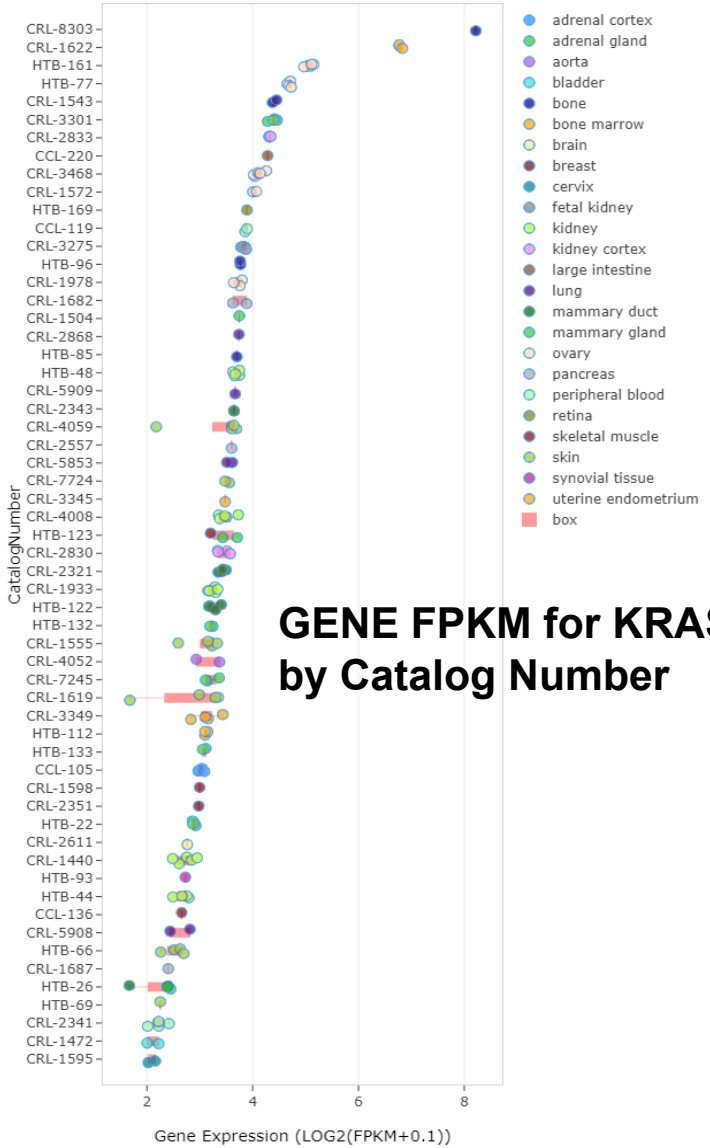
QC Metrics of ATCC® Cell Line Land

Quality Control Metrics	Average Values	Metadata Available
RNA Integrity Number (RIN)	>6.5	▪ Passage # ▪ Sex ▪ Race ▪ Age ▪ Disease ▪ Tissue/cell type ▪ Growth media ▪ Culture condition ▪ Cryopreservation
Nanodrop 260/280 Value	>1.8	
Input Sequence Read Number	$18x10^6$	
% Uniquely Mapped Reads	>80%	

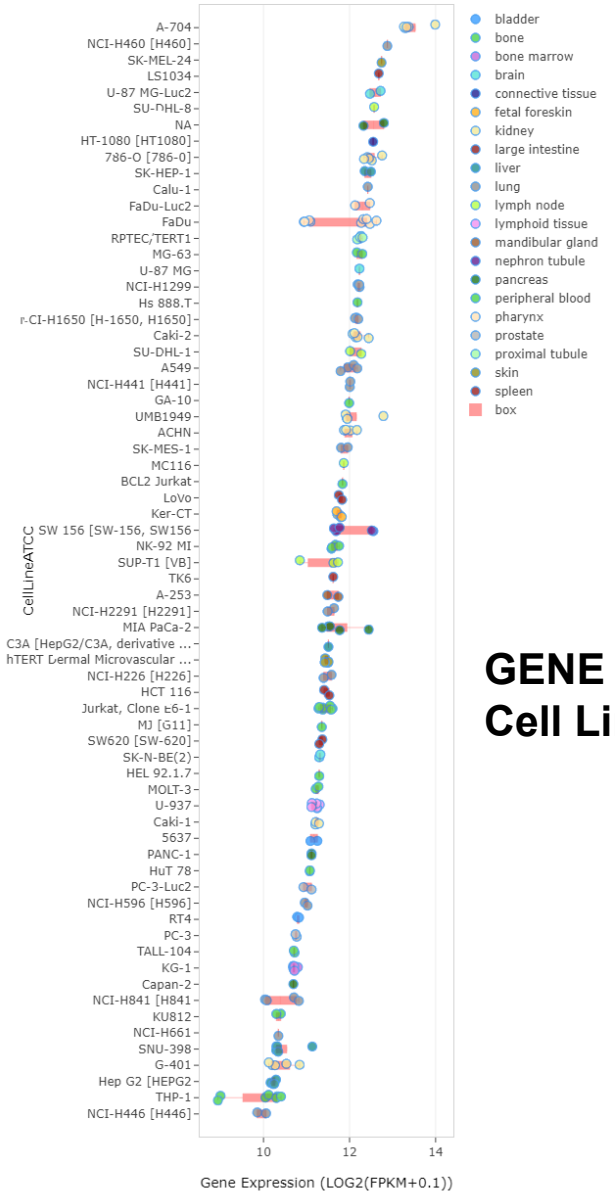
ATCC

ATCC® Cell Line Land – Example

A partnership with QIAGEN® Digital Insights



GENE FPKM for KRAS by Catalog Number



GENE FPKM for GAPDH by Cell Line Name (ATCC®)

ATCC® Cell Line Land – available through QIAGEN®

A partnership with QIAGEN® Digital Insights

QIAGEN | Solutions | Products and Services | Resources | Support | About | Request a Trial | Product Login | Contact | Shop

ATCC Cell Line Land

Manually curated cell line 'omics data from the most popular cell lines in ATCC's collection

ATCC Cell Line Land is a continually growing database of cell line 'omics data from both common and novel human and mouse cell lines and primary tissues and cells from ATCC. It empowers you to precisely plan and design your preclinical experiments by speeding up cell line characterization with unique, high-quality cell line 'omics data from a trusted source.

REQUEST A CONSULTATION

Currently includes **Authenticated RNAseq Data** for over 200 ATCC® cell lines.

https://digitalinsights.qiagen.com/atcc-cell-line-land/

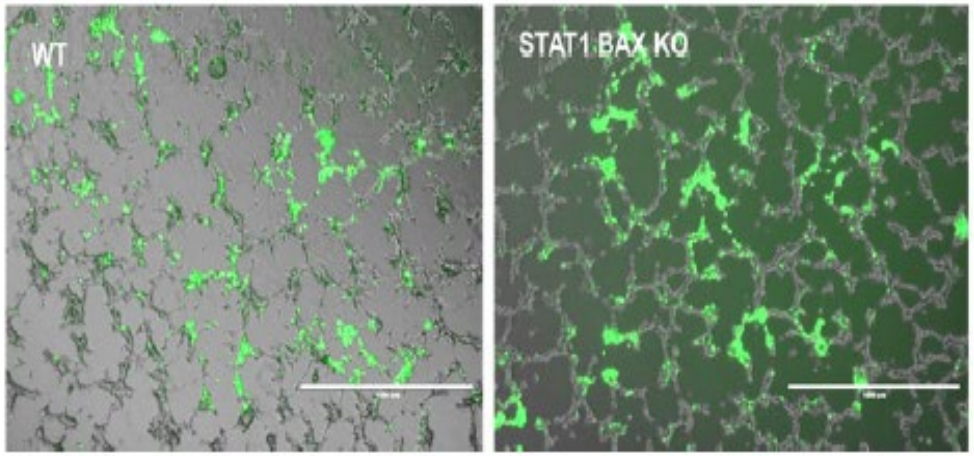
ATCC

Increased viral production in 293.STAT1 BAX KO cells

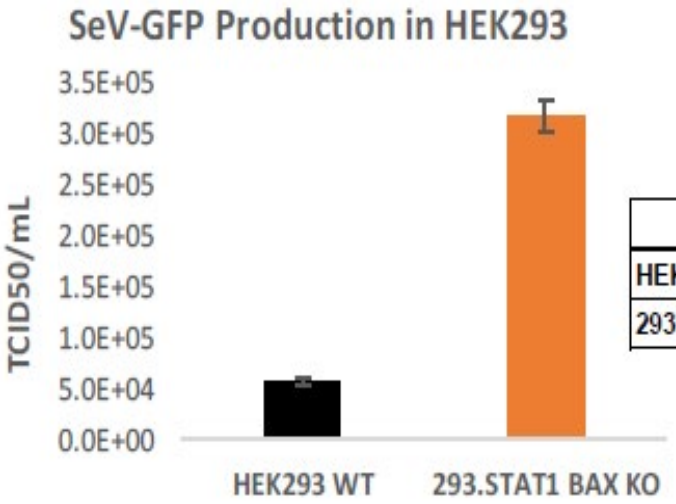
A

Dilution	HEK293 WT	293.STAT1 /BAX KO
-7	-	-
	-	-
	-	-
-6	-	-
	-	-
	-	+
-5	-	+
	+	-
	-	+
-4	+	+
	+	+
	+	+
-3	+	+
	+	+
	+	+
-2	+	+
	+	+
	+	+

B

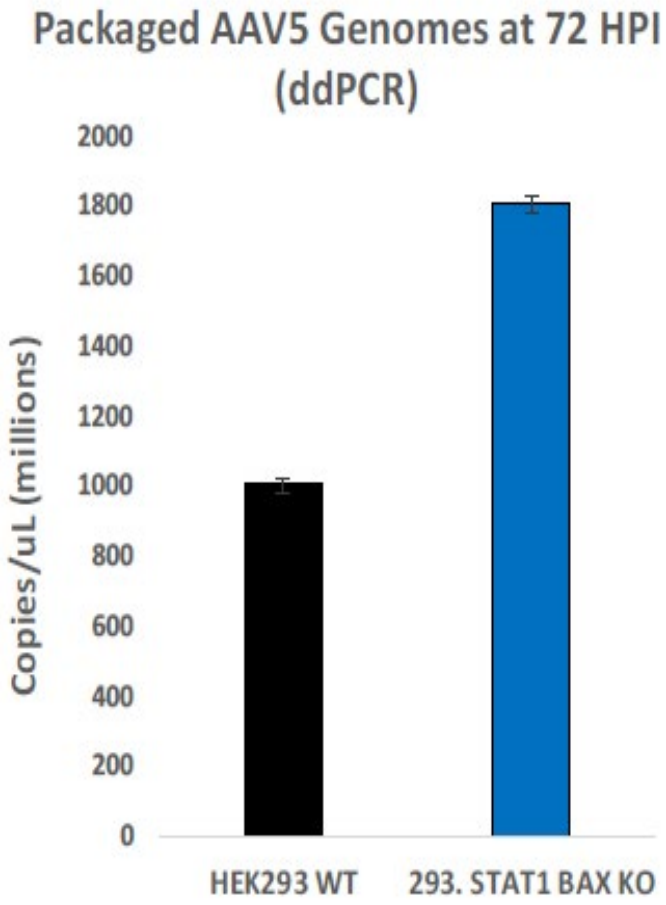


C



	TCID50/mL
HEK293 WT	5.62E+04
293.STAT1 BAX KO	3.16E+05

D



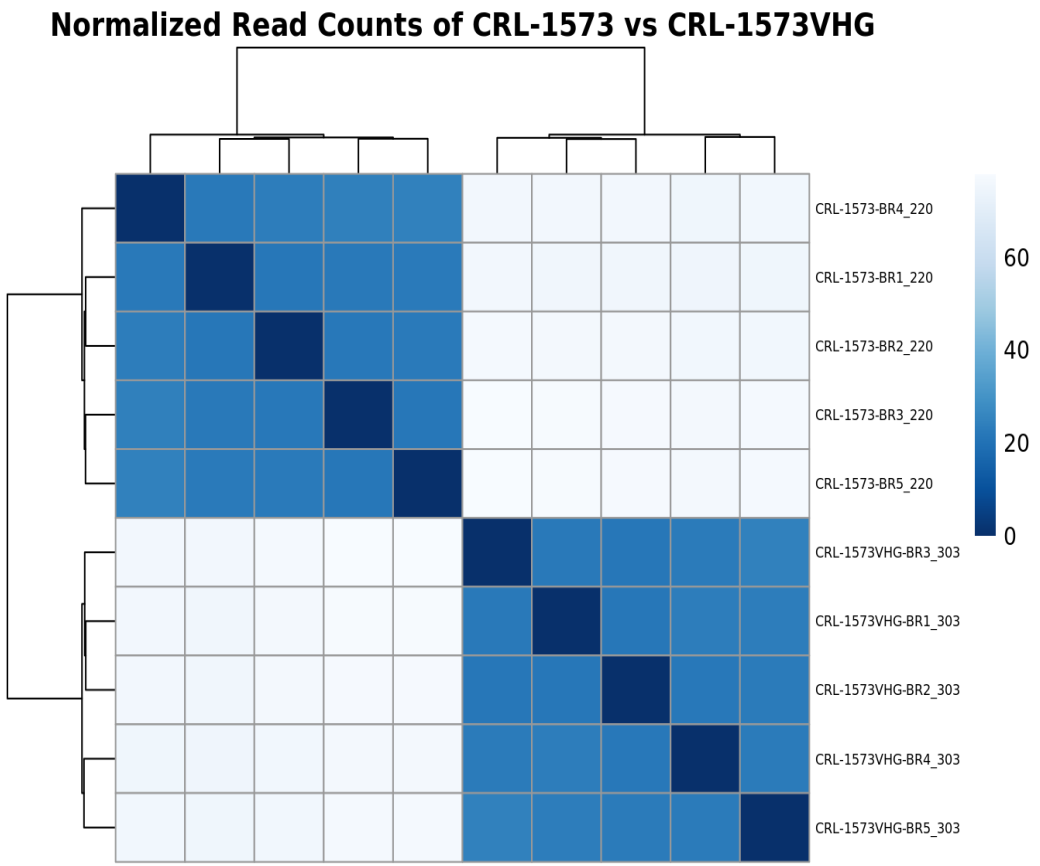
ATCC

Differential gene expression

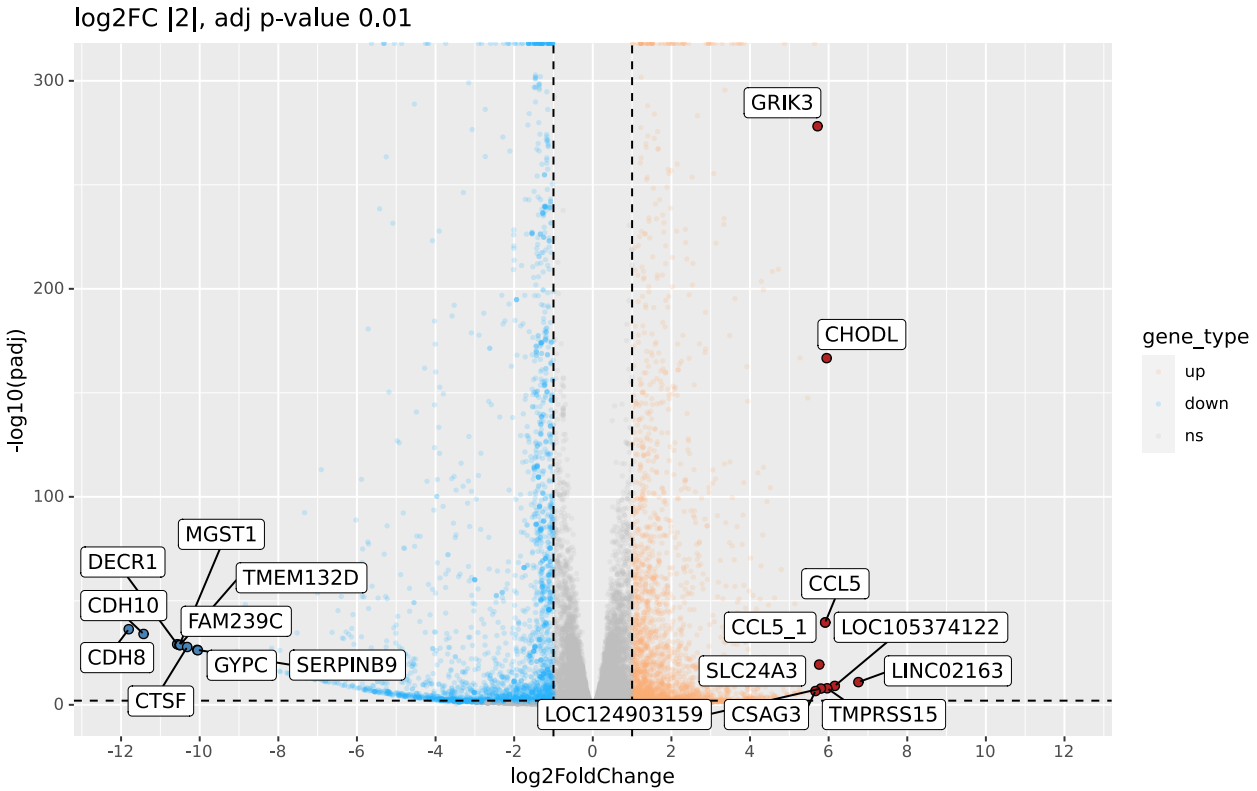
Distance correlation heatmap between normalized gene counts in 293 WT and 293.STAT1 BAX KO cells

Volcano plot showing top 20 differentially expressed genes in WT relative to KO cells

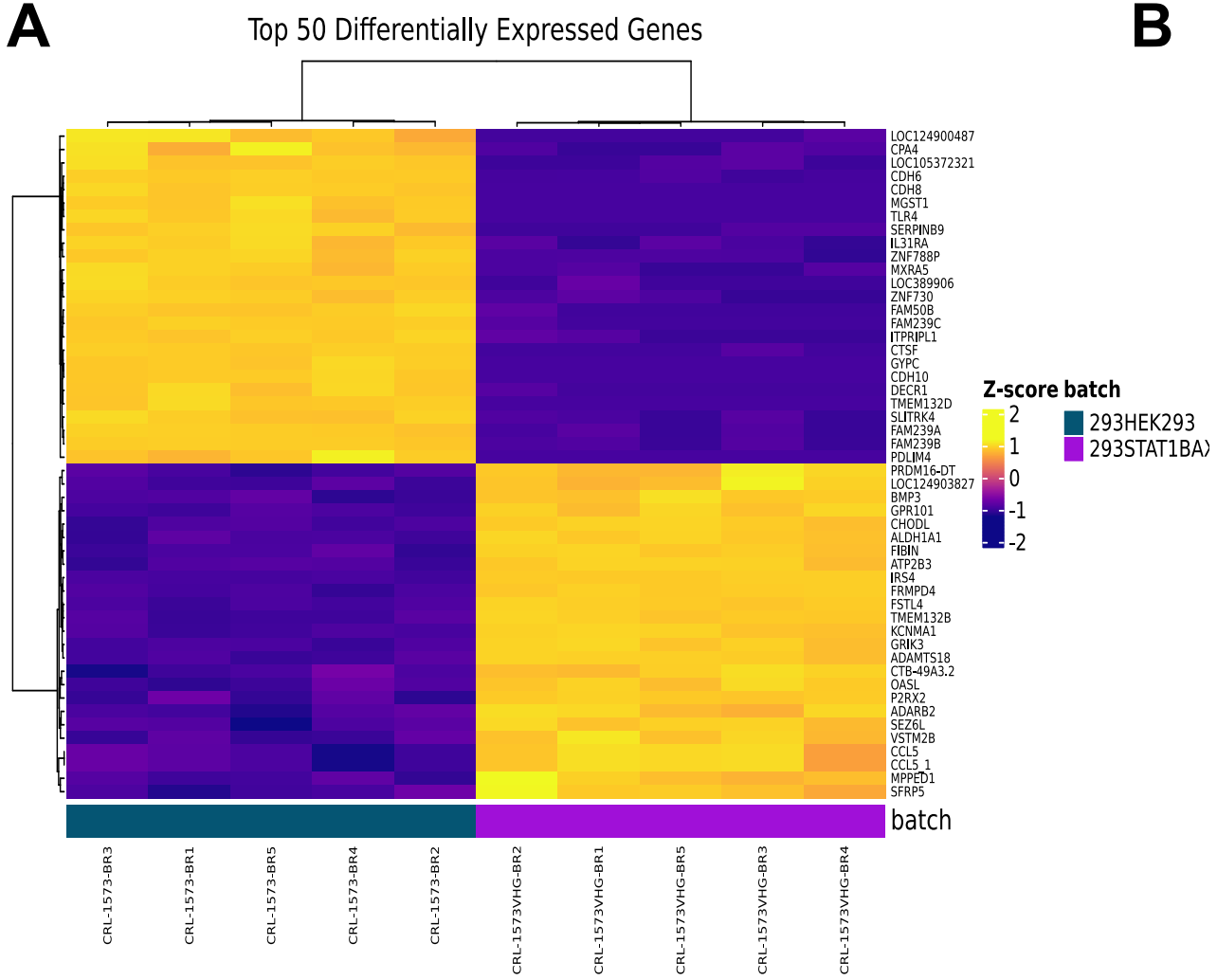
A

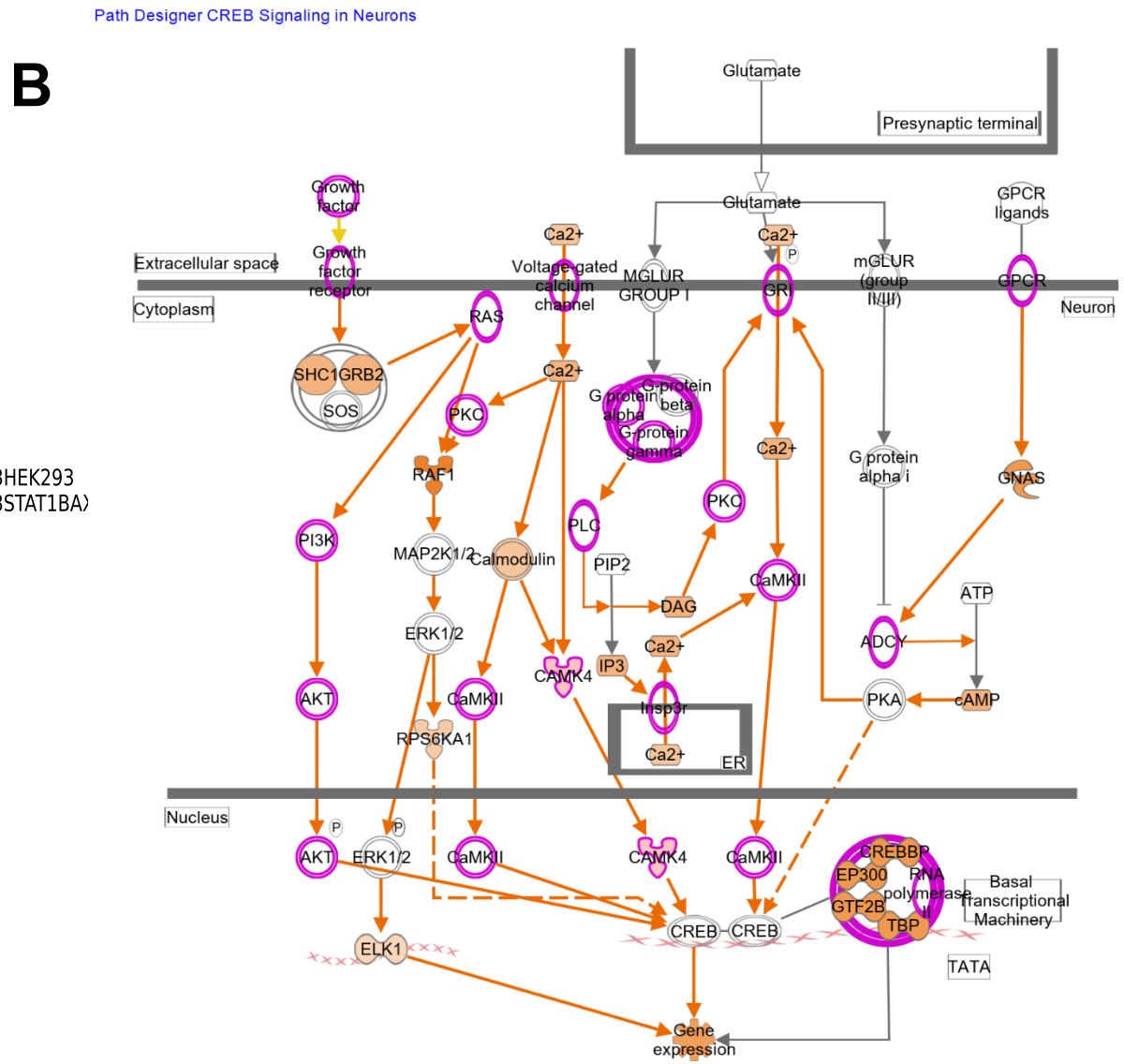


B



Pathways activated promote cell growth





Key takeaways

- ATCC® characterized the whole transcriptome of 70+ authenticated human mouse kidney cell lines, and 150+ cancer immune cell lines from the ATCC® biorepository.
- The data produced in this study are intended to be used as molecular reference standards and will be available to the scientific community within the ATCC® Cell Line Land.
- ATCC® Cell Line Land contains complete transcriptome data from new cell lines of different tissue and disease types, and 1000 samples will be added each year.

ATCC sequencing & bioinformatics center

Ajeet Singh, PhD
Senior Scientist, Bioinformatics BioNexus
asingh@atcc.org

Genomics Team

Briana Benton, PMP
Ana Fernandes
Stephen King, MSc
James Duncan, MSc
Samuel Greenfield, MSc
Corina Tabron, MSc
Noah Wax, MSc
Rula Khairi, MSc
Robert Marlow
Jade Kirkland

Bioinformatics Team

John Bagnoli
Scott Nguyen, PhD
David Yarmosh, MSc
Nikhita Putheveetil, MSc
P. Ford Combs, PhD
Amy Reese, MSc

ATCC | Credible leads to Incredible™

Group Lead

Jonathan Jacobs, PhD
Senior Director, Bioinformatics
BioNexus Principal Investigator
jjacobs@atcc.org
@bioinformer

Thank you!

ATCC Cell Line Land	https://digitalinsights.qiagen.com/atcc-cell-line-land/

ATCC