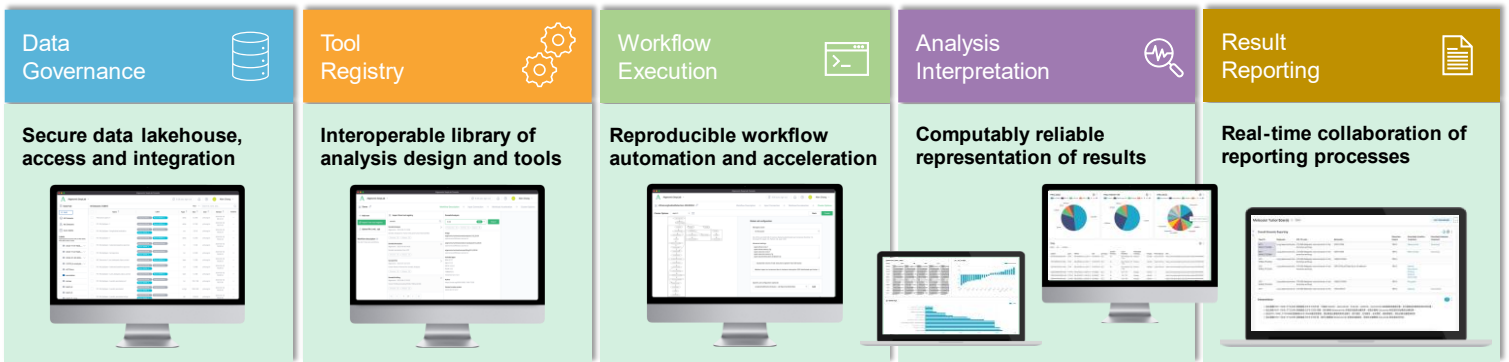




# Data and AI-driven Precision Medicine

Secure cloud-based data management and analysis workflow of genomic medicine

Atgenomix SeqsLab Platform governs the large volume of rapidly growing genomics and omics data from different layers of biological regulation and automates the diverse execution of commercially-available, open-source, and proprietary analysis workflows from sequencing to reporting at scale and speed.



## Audited quality

Comply with the most relevant frameworks and rigorous compliance standards in the healthcare industry: ISO/IEC 27001, ISO/IEC 27018, ISO 13485, IEC 62304, FDA/MDCG cybersecurity guidance, FDA 21 CFR Part 11 audit trail, GA4GH, GDPR, and more.

## Faster turnaround

Automate end-to-end workflows on all data with fully-managed cloud-native CPU/GPU parallel computing infrastructure, automatically scaling compute resources based on workload requirements to achieve operational efficiency.

## Customizable analysis

Build scalable and reproducible workflows for a wide range of analyses by combining WDL (Workflow Description Language), SQL (Structured Query Language), AI/ML, and GraphQL (Graph Query Language) into a unified workflow.

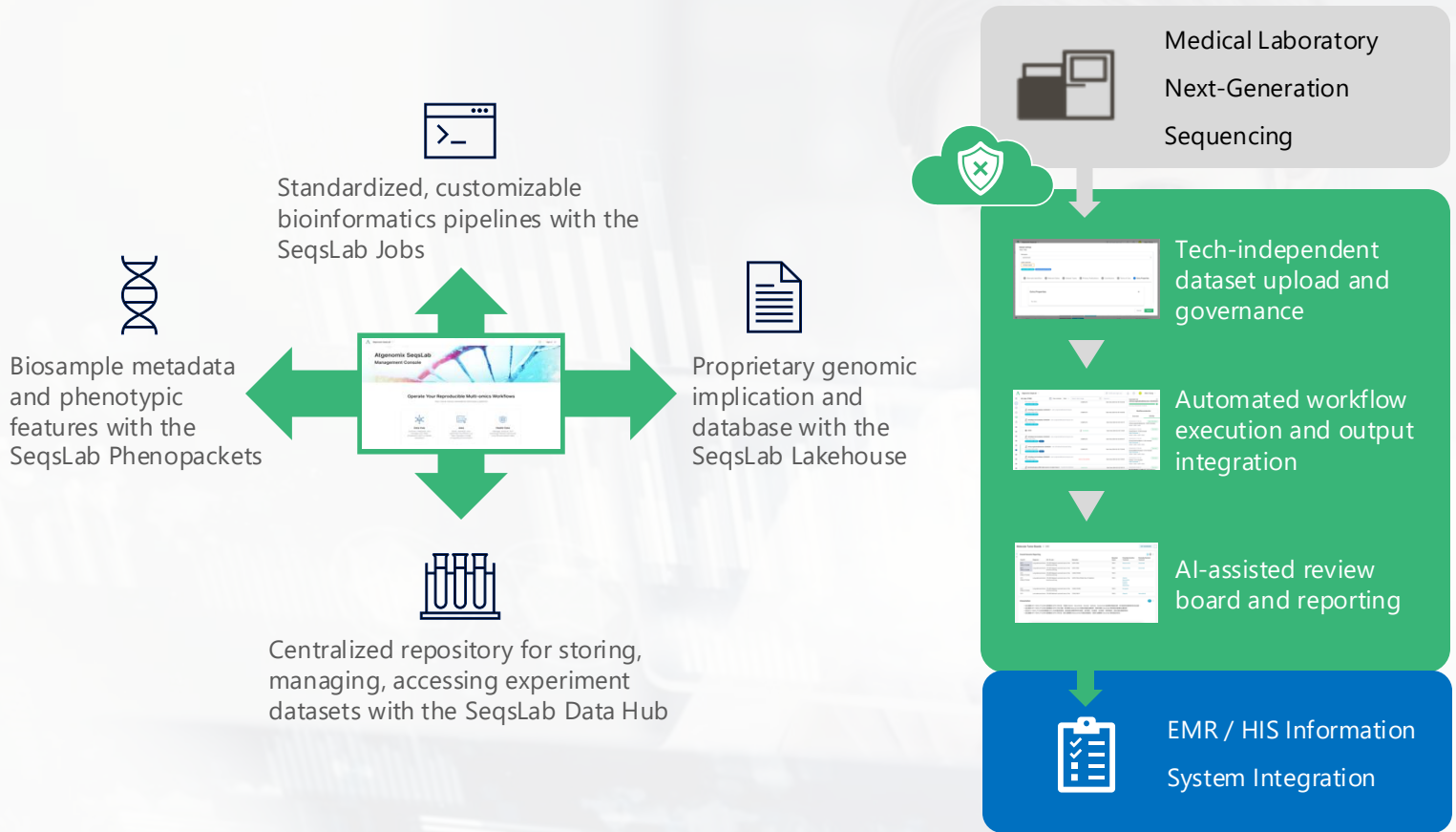
## Smarter usage

Optimize distributed workflow scheduling by partitioning datasets intelligently and leveraging in-memory processing capabilities to make efficient use of spot compute resources and to reduce the need for manual intervention.

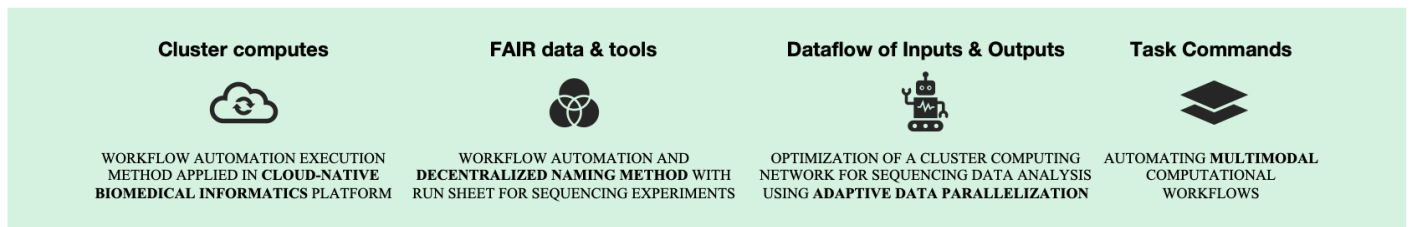
## Ready for routine clinical genomics

- ▶ Whole genome sequencing
- ▶ Whole exome sequencing
- ▶ Tumor-only and tumor-normal somatic analysis
- ▶ Longitudinal circulating tumor DNA profiling
- ▶ RNAseq, single-cell RNA analysis
- ▶ Multi-trait whole genome regression test
- ▶ SNP array analysis

# Integrating Atgenomix SeqsLab in your routine clinical genomics



## Exceptional Workflow Scalability

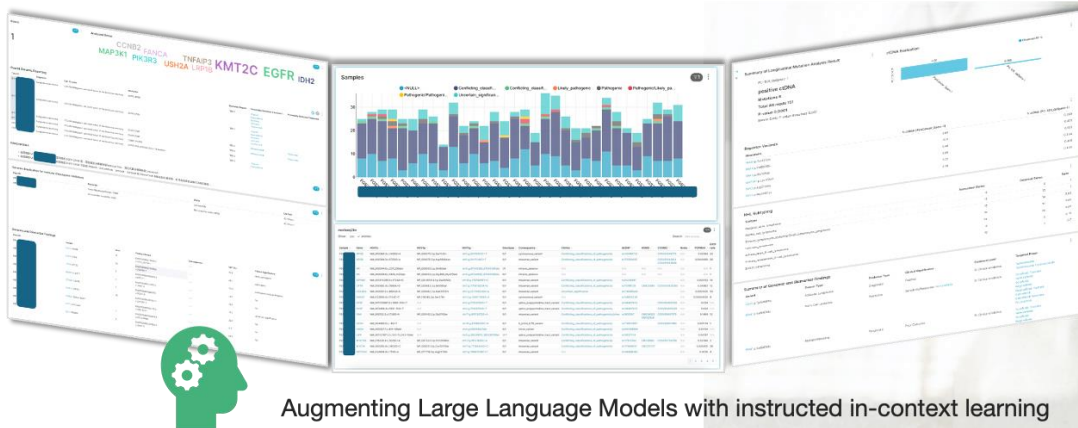


Workflow	Before	Increased throughput performance	Reduced compute cost <sup>(4)</sup>	Result consistency
Whole genome sequencing secondary and tertiary analysis <sup>(1)</sup>	> 25 hours	+10X (2.8 hours)	70%	99.9%
Joint genotyping of whole genome for 2,500 samples <sup>(2)</sup>	> 1 month	+16X (40 hours)	60%	100%
Whole genome regression of 10 traits for 100,000 samples <sup>(3)</sup>	> 90 hours	+20X (4.5 hours)	90%	100%

(1). Broad institute BWA-MEM/GATK4 best practice, including QC.  
 (2). Broad institute GATK v4.2 workflow.  
 (3). Regenie pipeline developed by Regeneron Genetics Center.  
 (4). The cost for the community workflows is based on the cost of running the workflows in a typical 64-core, 512GiB RAM Linux virtual machine (Azure E64d\_v5).

# Democratizing Access to Make Discoveries that Improve Healthcare Outcomes

Dedicated omics data lakehouse combined with Generative AI to identifying and prioritizing diagnostic targets in real-time and unleashing the power of the world's best scientific minds.



Augmenting Large Language Models with instructed in-context learning

- + Molecular tumor board
- + Precision health management
- + Differential diagnostics
- + Translational medicine research



## Robust Body of Evidence

**iScience**  
Volume 25, Issue 11, 21 October 2023, 100981

**Multicentric characterization and drug testing establish circulating tumor cells as an ex vivo tool for personalized medicine**

Jin Tang Chen<sup>1,2,3,4</sup>, Han-Hsin Chen<sup>1,2,3</sup>, Jian-Chuan Liu<sup>1,2,3</sup>, Ting-Ying Huang<sup>1,2,3</sup>, Jian-Chao Liu<sup>1,2,3</sup>, Chia-Chun Guo<sup>1,2,3</sup>, Chen-Hsiang Tang<sup>1,2,3</sup>, Chung-Tai Li<sup>1,2,3</sup>, Jui-Yen Wang<sup>1,2,3</sup>, Edward Lu<sup>1,2,3</sup>, Hsin-Fan Liu<sup>1,2,3</sup>, Yu-Ying Wu<sup>1,2,3</sup>, Chia-Hsiang Tang<sup>1,2,3</sup>, Alexander B. Dunn<sup>1,2,3</sup>, Carlo W. Tsai<sup>1,2,3</sup>, Yu-Chen Wang<sup>1,2,3,5,6,7,8,9</sup>, Ji-Bin Alsharif Tang<sup>1,2,3,4,9,10,11,12</sup>, J. B. Alsharif Tang<sup>1,2,3,4,9,10,11,12</sup>, J. B. Alsharif Tang<sup>1,2,3,4,9,10,11,12</sup>

**Abstract**  
Current human genome sequencing assays in both clinical and research settings primarily utilize short-read sequencing and apply resequencing pipelines to detect genetic variants. However, these resequencing-based assays pipelines remain a considerable challenge due to an incomplete reference genome, mapping errors and high sequence divergence. To overcome this challenge, we propose an efficient and effective whole-read assembly workflow with unannotated graph mining algorithms on an Apache Spark large-scale data processing platform called ConnectReads. It fully utilizes short-read data information. ConnectReads is able to generate assembled contigs and then identify de novo mutations. It provides high-resolution SV discovery than that provided by other methods, especially a high diversity against reference and in-gap regions of reference. Furthermore, we demonstrate a cost-effective approach by leveraging ConnectReads to investigate all species of genetic changes in population-scale studies.

**phSeq: Accelerating String Graph Construction I Novo Assembly on Spark**

Tai-Sai Su<sup>1,2</sup>, Ming-Tai Chang<sup>1</sup>, Yan-Chian Cheng<sup>1</sup>, Yun-Lung Li<sup>1</sup> and Yao-Ting Wang<sup>1</sup>

**Abstract**  
De novo genome assembly is an important application on both underlabeled genome assembly and variation in a reference-guided assay. In comparison with de Bruijn graph, string graph is a better data structure for de novo assembly. However, string graph construction is computational intensive. We propose Graph-based String Graph Construction (phSeq) to accelerate string graph construction by leveraging the distributed computing framework. phSeq is implemented with Scala on Spark and they avoid the expensive graph construction process. phSeq is implemented with Scala on Spark and they avoid the expensive graph construction process. Supplementary data are available at Bioinformatics online.

**Genome analysis SeqLab: an integrated platform for cohort-based annotation and interpretation of genetic variants on Spark**

Ming-Tai Chang<sup>1,2</sup>, Yan-Chian Cheng<sup>1</sup>, Jian-Chao Liu<sup>1,2</sup>, Han-Hsiang Tang<sup>1,2</sup>, Yu-Lung Li<sup>1,2</sup>, Yao-Ting Wang<sup>1,2</sup>, Chia-Yi Chen<sup>1,2,3</sup> and Chung-Tai Li<sup>1,2,3</sup>

**Abstract**  
SeqLab is a platform that helps researchers to easily annotate and interpret genetic variants. It offers a large quantity of personal genomes, a pipeline to integrate multiple data to provide the variants based on variant distributions and an on-site annotation on the effects of the variants. SeqLab also offers the variant effect interpretation, such as the interpretation of the variant effect on the basis of the process of alternative splicing and transcription. The key features of SeqLab are variant annotation on large structural variations, diverse combinations of variant files, easy integration with a wide amount of public databases, and available visualization of variant effects. SeqLab is implemented with Scala on Spark and they avoid the expensive graph construction process. Supplementary data are available at Bioinformatics online.

**Detection of Rare Methyl-CpG Binding Protein 2 Gene Missense Mutations in Patients With Schizophrenia**

Jian-Chao Liu<sup>1,2,3,4</sup>, Jian-Chuan Liu<sup>1,2,3,4</sup>, Ting-Ying Huang<sup>1,2,3,4</sup>, Jian-Chao Liu<sup>1,2,3,4</sup>, Chia-Chun Guo<sup>1,2,3,4</sup>, Chen-Hsiang Tang<sup>1,2,3,4</sup>, Chung-Tai Li<sup>1,2,3,4</sup>, Jui-Yen Wang<sup>1,2,3,4</sup>, Edward Lu<sup>1,2,3,4</sup>, Hsin-Fan Liu<sup>1,2,3,4</sup>, Yu-Ying Wu<sup>1,2,3,4</sup>, Chia-Hsiang Tang<sup>1,2,3,4</sup>, Alexander B. Dunn<sup>1,2,3,4</sup>, Carlo W. Tsai<sup>1,2,3,4</sup>, Yu-Chen Wang<sup>1,2,3,4,5,6,7,8,9</sup>, Ji-Bin Alsharif Tang<sup>1,2,3,4,5,6,7,8,9</sup>, J. B. Alsharif Tang<sup>1,2,3,4,5,6,7,8,9</sup>

**Abstract**  
Methyl-CpG binding protein 2 (MeCP2) is a DNA-binding protein that recognizes a specific DNA sequence (5'-CCG-3') and binds to it. MeCP2 is a transcription factor that regulates gene expression. MeCP2 is also a tumor suppressor protein. MeCP2 mutations have been found in patients with schizophrenia. We used a whole-genome sequencing approach to identify rare MeCP2 missense mutations in patients with schizophrenia. We found 17 rare MeCP2 missense mutations in 17 patients. These mutations were associated with schizophrenia. We found that MeCP2 mutations were associated with schizophrenia. We found that MeCP2 mutations were associated with schizophrenia.

**Copy number variant hotspots in Han Taiwanese population induced pluripotent stem cell lines - lessons from establishing the Taiwan human disease iPSC Consortium Bank**

Chia-Hsiang Tang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Ting-Ying Huang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Jian-Chao Liu<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Chia-Chun Guo<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Chen-Hsiang Tang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Chung-Tai Li<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Jui-Yen Wang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Edward Lu<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Hsin-Fan Liu<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Yu-Ying Wu<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Chia-Hsiang Tang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Alexander B. Dunn<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Carlo W. Tsai<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Yu-Chen Wang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Ji-Bin Alsharif Tang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, J. B. Alsharif Tang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>

**Abstract**  
Copy number variants (CNVs) are structural variations that affect the number of copies of a particular DNA segment. CNVs are common in the human genome and have been found to be associated with a variety of human diseases. We used a whole-genome sequencing approach to identify CNVs in Han Taiwanese population induced pluripotent stem cell lines. We found 17 rare CNVs in 17 patients. These CNVs were associated with schizophrenia. We found that CNVs were associated with schizophrenia.

**Identification of Rare Mutations of Two Presynaptic Cytomatrixes RSN and PCL0 in Schizophrenia and Bipolar Disorder**

Yu-Lung Li<sup>1,2</sup>, Ming-Tai Chang<sup>1,2</sup>, Yan-Chian Cheng<sup>1,2</sup>, Chung-Tai Li<sup>1,2,3</sup>, Chia-Yi Chen<sup>1,2,3,4</sup> and Yao-Ting Wang<sup>1,2,3,4</sup>

**Abstract**  
RSN and PCL0 are two presynaptic cytomatrixes that are involved in the formation of presynaptic cytomatrixes. We used a whole-genome sequencing approach to identify rare mutations of RSN and PCL0 in patients with schizophrenia and bipolar disorder. We found 17 rare mutations in 17 patients. These mutations were associated with schizophrenia and bipolar disorder. We found that mutations were associated with schizophrenia and bipolar disorder.

**Identification of a novel nonsense homozygous mutation of LINS1 gene in two sisters with intellectual disability, schizophrenia, and anxiety**

Chia-Hsiang Tang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Ting-Ying Huang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Jian-Chao Liu<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Chia-Chun Guo<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Chen-Hsiang Tang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Chung-Tai Li<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Jui-Yen Wang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Edward Lu<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Hsin-Fan Liu<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Yu-Ying Wu<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Chia-Hsiang Tang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Alexander B. Dunn<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Carlo W. Tsai<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Yu-Chen Wang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Ji-Bin Alsharif Tang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, J. B. Alsharif Tang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>

**Abstract**  
LINS1 is a protein that is involved in the formation of presynaptic cytomatrixes. We used a whole-genome sequencing approach to identify a novel nonsense homozygous mutation of LINS1 in two sisters with intellectual disability, schizophrenia, and anxiety. We found 17 rare mutations in 17 patients. These mutations were associated with intellectual disability, schizophrenia, and anxiety.

**Rare Mutations of SCN9A, DPP4, ABCA13, and ANLN in Bipolar Disorder**

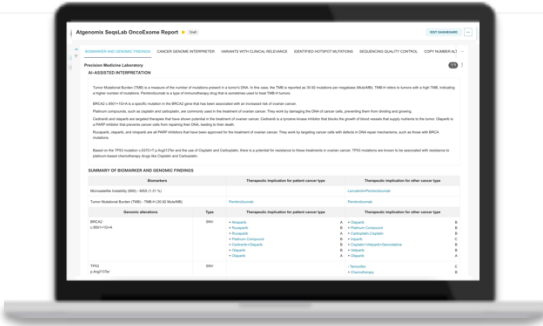
Chia-Hsiang Tang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Ting-Ying Huang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Jian-Chao Liu<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Chia-Chun Guo<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Chen-Hsiang Tang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Chung-Tai Li<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Jui-Yen Wang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Edward Lu<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Hsin-Fan Liu<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Yu-Ying Wu<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Chia-Hsiang Tang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Alexander B. Dunn<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Carlo W. Tsai<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Yu-Chen Wang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, Ji-Bin Alsharif Tang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, J. B. Alsharif Tang<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>

**Abstract**  
SCN9A, DPP4, ABCA13, and ANLN are genes that are involved in the formation of presynaptic cytomatrixes. We used a whole-genome sequencing approach to identify rare mutations of SCN9A, DPP4, ABCA13, and ANLN in patients with bipolar disorder. We found 17 rare mutations in 17 patients. These mutations were associated with bipolar disorder. We found that mutations were associated with bipolar disorder.





# Atgenomix SeqsLab OncoExome Whole Exome Analytical Solution for Cancer



The genomic profiling application that integrates capture-based target enrichment kits with the data analytics capabilities and accelerated computing features of Atgenomix SeqsLab Platform.

Atgenomix SeqsLab OncoExome provides standardized detection workflow and enhanced analytical capabilities for the comprehensive cancer genomic profiling.

## Main Features

Comprehensive coverage of multiple types of variants, and actionable genomic biomarkers in up to 1,275 genes, enabling data-guided decision making.

Exome Library	Variants Called	Biomarkers Analyzed	QC Metrics
<ul style="list-style-type: none"> <li>Capture-based target enrichment</li> <li>Sequencing with Unique Molecular Identifier (UMI)</li> </ul>	SNVs Indels CNVs *	MSI TMB LOH * HRD * Splicing prediction Therapy implications	Total bases $Q \geq 30$ Duplicate reads Mapping rate On-target reads Coverage of targeted regions Coverage uniformity

\* An intended Panel of Normals (PON) with similar technical properties of the tumor is required for the analyses. Atgenomix provides additional workflows for generating PON.

## Analytical Capabilities

SeqsLab OncoExome analyzes complex WES data by calling, annotating and classifying genomic variants in all the targeted regions.

Turnaround time from FASTQ	~3 hours
Tumor-only analysis	Available
Genome Reference	GRCh38

Optimized table output for easy interpretation to genes, alterations, and biomarkers of interest.

## Integrated Workflow for Genomic Analysis, Interpretation and Reporting

