

# Literature Lab<sup>TM</sup> analysis of neurodevelopmental pathways in a stem cell model of Schizophrenia

Damon Anderson, PhD, Applications Scientist and Paul Martinez, President and CEO, Acumenta Biotech, Westminster, MA. Co-responding author: Damon Anderson, PhD, danderson@acumenta.com, 412-901-7785

Genetic factors play a prominent role in Schizophrenia (SZ) susceptibility and modern large cohort studies combined with high-resolution mapping techniques have revealed many significant genetic risk factors. Animal models and postmortem brain tissue have served as important systems in the study of SZ, however, inherent limitations include the inability to recapitulate neurodevelopmental aspects of the disease among other issues. Human induced pluripotent stem cell (hiPSC)-derived neural progenitor cells (NPCs) have significant promise in this regard, and recent studies have sought to characterize NPCs to explore the developmental pathways of SZ and the use of this system for diagnostic and therapeutic discovery. In this work, we employed the Literature Lab™ PLUS platform to analyze recently published microarray gene expression data from SZ patient hiPSC-derived NPCs compared with controls. Literature Lab<sup>TM</sup> PLUS identified several pathways in common with Gene Ontology (GO) analysis used in the published work, including cell remodeling and oxidative stress. Interestingly, Literature Lab<sup>™</sup> PLUS also identified pathways involving axon guidance, synapse formation, neuronal maturation, long term potentiation, and others that have been previously limited to SZ end point studies using post mortem brain tissue. Consistent with the exploratory power of the technology, Literature Lab™ PLUS identified additional emerging pathways, several of which have been linked to other neurodevelopmental and neurodegenerative diseases such as Alzheimer's and Bipolar Disorder. This work highlights the capabilities of Literature Lab<sup>™</sup> in identifying novel pathway associations, thereby broadening our understanding of the molecular pathogenesis of SZ and other neuropsychiatric diseases.

Schizophrenia is a severe psychiatric disorder with a world-wide prevalence of 1 %. The disease is characterized by the presence of delusions, hallucinations, disorganized speech or behavior, and a host of other symptoms. Although the characteristic symptoms of schizophrenia (SZ) generally appear late in adolescence, it is thought to be a neurodevelopmental condition that can appear in early life. Postmortem studies have identified defects in neurons, such as reduced neuronal size and spine density in the prefrontal cortex and hippocampus, however, these findings represent disease end points. There is a need for reliable models for neurodevelopmental studies in order to understand SZ predisposition and initiation. Based on existing models and genetic data, one prevailing hypothesis is that disruptions in fetal prefrontal cortical development underlie SZ.

Animal models and post mortem brain samples have provided valuable insight into the roles of specific genes and pathways, and the pathophysiology of SZ. However, there are limitations including: lack of accuracy in model human brain complexity, and post mortem degradation in the fidelity of the samples, among other issues. Moreover, these approaches have had limited success in understanding the development, initiation, and spatial temporal progression of the disease.

Stem cell model systems have been recently developed to address the major challenges of studying the underlying genetic mechanisms and the developmental pathways of SZ. These live cell models have been established by reprogramming patient somatic cells, such as fibroblasts, into human induced pluripotent stem cells (hiPSCs), which can then be coaxed to differentiate into neural progenitor cells and neurons. hiPSCs, and their differentiated neural progenitor cells (NPCs) and neurons, possess the complex genetic backgrounds that



exist in patients. They have the potential to provide critical new insights into the neurodevelopmental basis of disease and may prove useful in stratifying targets of diagnostic and therapeutic development.

A recent study explored the use of patient based hiPSC-derived neural progenitor cells to study gene expression and the underlying developmental mechanisms of SZ. *Brennand et al. 2015.* Phenotypic differences in hiPSC NPCs derived from patients with schizophrenia. *Molecular Psychiatry. 20: 361-368.* The authors first characterized these NPCs as valid models of fetal brain tissue and neurodevelopment by comparing gene expression of hiPSC-derived NPCs and neurons to the Allen BrainSpan Atlas. They then performed microarray analysis and Gene Ontology (GO) pathway annotation of SZ cells versus controls, followed protein expression and phenotypic analysis.

Phenotypic analyses of SZ patient NPCs showed aberrant migration and reduced chaining compared to control NPCs, suggesting reduced expression of cell migration and synapse formation genes in SZ NPCs consistent with a loss of neuron connectivity. Microarray gene expression analysis was performed using Affymetrix Human 1.0ST arrays and the Partek Genomics suite software, followed by Gene Ontology (GO) pathway analysis based on manually curated database information.



Shown in Supplementary Information Figure 4B. are hierarchical clustering and heat maps of control and SZ hiPSC NPCs of 487 unique genes (135 up-regulated and 352 down-regulated) with greater than 1.30-fold expression changes (p<0.05). Shown in C. are heat maps of altered expression of GO pathway cellular adhesion and extracellular matrix genes in SZ NPCs. For expression microarrays, three independent NPC samples for each of the four SZ patients as well as four control subjects were compared.

The results indicate that SZ NPCs exhibit abnormal cell remodeling and oxidative stress in addition to perturbations in synapse formation and neuronal connectivity. The conclusions have obvious impact on the further characterization of SZ NPCs and insight into the neurodevelopmental pathways underlying the disease. However, Gene Ontology (GO) pathway analysis is inherently limited in its treatment of the expression data set and interrogation of pathway information. The GO annotation database is based on annotations by manual curation of current literature. There is always an unknown duration of time lag necessary for the manual curation of the literature and thus, the database is never up to date.

Mol Psychiatry. 2015 Mar;20(3):361-8. doi: 10.1038/mp.2014.22.

© 2016 Acumenta Biotech www.acumenta.com



Literature Lab<sup>TM</sup> from Acumenta Biotech is a powerful alternative approach that involves analysis of the microarray data as a unique gene set, not a static entity, and direct statistical mining of the PubMed database, not simply a manually curated database. Visit <u>www.acumenta.com</u> or view an <u>application note</u> or a <u>peer reviewed study</u> to learn more about how Literature Lab<sup>TM</sup> works.

We enlisted Literature Lab<sup>TM</sup> PLUS in the analysis of the microarray data set from the published work. The 135-gene upregulated set and the 352-gene downregulated set were analyzed separately.

-Lit <u>eratu</u> re			E	Based on PubMed abstracts from (	PSC NPCS SZ VS C V3_OP 01/01/1990 through 02/28/2015			-
domain name or term to view t	he details.	Click on an	association	count to display the list of tern	ns. Right Click for Annotations.			
Domai	ns & Associ	ation Cou	ints		Strong	Assocations for		
	Strong	Moderat	te Positive	Visualizer	Term		P-Value	-
Pathways	<u>3</u>	9 5	<u> </u>	Res .	Cell Adhesion	2.77	0.0028	
Disassas MaSH		0.0	20	0	Mitochondria Apoptotic	2.43	0.0076	
Diseases-riesii	<b>* *</b>	2	<b>20</b>		Hematopoietic Cell Lineage	2.36	0.0091	-
PathologicalCond-MeSH	0	0	2		Moderate	Assocations for		
PharmacoAction-MeSH	0	2	8	Re-	P	athways		
				-	Term	Score 🕷	P-Value	
PharmacoSubst-MeSH	<b>1</b>	9 4	<u> </u>		Cell Adhesion Molecules	2.72	0.0033	-
ChemicalActions-MeSH	01	3	13		PDGF	2.22	0.0133	
ChemicalsDrugs-MeSH		20	174	R	Apoptosis	1.96	0.0249	
chemicalsor ugs Tresh	<u> </u>	<b>2</b>	121		Free Radical Apoptosis	1.92	0.0273	
Anatomy-MeSH	0	<u>6</u>	9 9	<b>Nor</b>	IL 3	1.67	0.0478	-
Physiology-MeSH	0	0	2	<b>*</b>	Positive	Assocations for athways		
CellPhysiology-MeSH	1	3	<u> 3</u>		Term	Score V	P-Value	
CellTypes-MeSH	0	4	9	<b>~</b>	FOXO Family	4,43	0.0000	-
e lles a su eu				-	EPHA	3.74	0.0001	
CellStructures-MeSH	1	2	<b>9</b> 12		C-KIT	3.46	0.0003	
CellLines	0	2	3		Syndecan	3.30	0.0005	
TissueTypes-MeEH			2 2	8	Lymphocyte Diapedesis	2.67	0.0038	
<u>Hissuerypes Hesh</u>	<b>* *</b>	<b>Y</b>	<b>V E</b>		Lymphocyte Adhesion	2.67	0.0038	
Metabolism-MeSH	91	2	0		Caspase Cascade	2.31	0.0104	
Biogenetics-MeSH	0	<b>0</b> 1	4		Sphingeligid Riscuptherin	2.20	0.0112	
					NDE2	2.00	0.0198	
OtherBiology-MeSH	1	94	<u> 5</u>		BAD Phosphorylation	1.97	0.0246	
Psychology-MeSH	0	01	4	Re-	Encal Adhesion	1.95	0.0257	
					Sulfation	1.87	0.0305	
					Hematopoiesis	1.81	0.0351	
					DNA Damage	1.74	0.0414	
					Etoposide	1.65	0.0497	
					Integrin Cell Adhesion	1.63	0.0517	

Literature Lab<sup>™</sup> PLUS analysis Figure 1.A.

The 135-gene upregulated set demonstrated three strong, five moderate, and eighteen positive pathway associations. The cell adhesion, mitochondrial and free radical apoptosis pathway associations are congruent with the GO analysis results in the paper which also identified cell adhesion and oxidative stress pathways.

Several unanticipated associations were identified including the Edh/ephrin signaling pathway, which has been implicated in the regulation of a host of processes critical to embryonic development including axon guidance, formation of tissue boundaries, cell migration, and segmentation. Eph/ephrin signaling has recently been identified to play a critical role in the maintenance of several processes during adulthood including long-term potentiation, angiogenesis, and cancer.

GSK-3 is a central intracellular signaling pathway that is involved in cellular proliferation, migration, glucose regulation, inflammation, and apoptosis. The pathway has recently been the subject of much research as it has been implicated in a number of diseases including: Type II diabetes (Diabetes mellitus type 2), Alzheimer's disease, bipolar disorder, and cancer.



Clicking on the pathways domain associations hyperlink in the table above displays the data table below with: The associations ranked according to strength across the x-axis, the statististical data underlying the associations at the top of the y-axis, and the individual genes in the gene set and their relative contribution the the log of the product of frequency (LPF) in the lower portion of the y-axis. A detailed description of this information and the Literature Lab<sup>TM</sup> association scoring process can be found here <u>www.acumenta.com</u> and <u>here.</u>

⊳titer				Domain: Pathways Experiment: Gene list iPSC NPCs SZ vs C V3_Up Based on PubMed abstracts from 01/01/1990 through 02/28/2015							ui 49 (?) N T 7				
Click on column hea Percentages = prop	ders or buttor portion of Expe Clear/Set/Use	<b>rimen</b> Check	ort columns of It Set LPF cont Marks	: columns or rows. Click again to reverse sort order. Enter the first few letters of a gene and/or term to find a cell iet LPF contributed by a gene. Click on a cell to view the abstracts for the gene and term. Right Click for Annotation rks Create a Subset of the Domain Visualizer						to find a cell. for Annotation	Sins. Gene: Term: Find				
Pathways	Count	~	Cell Adhesion	Mitochondria Apoptotic	Hematopoietic Cell Lineage	Cell Adhesion Molecules	PDGF	Apoptosis	Free Radical Apoptosis	IL 3	FOXO Family	EPHA	C-KIT	Syndecan	
/iew All Genes For Term														*	
ssociation 🕨	_		Strong	Strong	Strong	Moderate	Moderate	Moderate	Moderate	Moderate	Positive	Positive	Positive	Positive	
ene x Term Abstracts	-	+	12110	3531	554	11476	334	13326	2191	398	2158	451	612	31	
erm Abstracts	-	+	/6866	18877	8268	56938	/749	188233	24202	5128	3942	815	4762	172	
onzero Genes	-		68	31	32	60	31	83	36	21	26	13	28		
r Inden Seke 🗖	Count		-1.23	-1.24	-2.33	-1.09	-2,12	-1.35	-1.98	-2.1/	-0.39	-0.43	-1.51	-1	
andom sets 😐	Count		1000	1000	1000	1000	1000	1000	1000	1000	1000	9999	1000	100	
xpeninent set	Score		2 77	2.43	27	2 72	72	1.06	1.07	1 67	4 42	2 74	2 46	2 2 2	
	D-Value		0.0028	0.0076	0.0091	0.0033	0.0133	0.0249	0.0273	0.0478	0,000	0.0001	0.0003	0.000	
	Score rank		0.0020	0.0070	15	0.0055	0.0133	20	33	51	0.0000	0.0001	0.0003	0.000	
vmbol/OrigID	May PE		90 28%	97 37%	92 30%	98.07%	96 92%	86.97%	83 36%	98 27%	99 92%	99 98%	99.05%	99.029	
CAM1	2394	3 🗐	90.28%	0.00%	0.86%	98.07% K	0,19%	0.54%	0.35%	0.87%	0.00%	0.00%	0.03%	0.019	
ASP9	972	1	0.05%	97.37%	0.31%	0.01%	0.01%	86,97%	83.36%	0.04%	0.01%	0.00%	0.00%	0.00	
DA	549	2	0.00%	0.00%	0.25%	0.00%	0.02%	0.02%	0.00%	0.02%		- 10	0.00%		
CND2	339	8	0.05%	0.02%	0.48%	0.01%	0.61%	1.38%	0.06%	0.19%	0.01%	0.00%	0.02%	0.00	
Q01	314	4	0.00%	0.02%	0.02%	0.00%	0.00%	0.20%	2.26%	0.01%	0.00%		-		
F1A1	294	8	0.00%	0.00%	0.00%	0.00%		0.02%	0.01%	0.00%	-	0.00%	0.00%		
CL	279	8	7.75% @	0.00%	0.01%	0.81%	0.17%	0.04%	0.00%	0.00%		0.00%	0.00%	0.05	
DXO1	278	8	0.02%	0.04%	0.09%	0.01%	0.10%	1.57%	2.04%	0.05%	99.92%	0.00%	0.00%	0.00	
FPI	217	4	0.05%	0.00%	0.03%	0.04%	0.02%	0.03%	0.00%	-	-		-	0.07%	
AST(GID:831)/**CAST	202	3	0.04%	0.14%	0.02%	0.00%	0.01%	0.61%	0.25%		0.00%				
ITLG	184	0	0.07%	0.00%	92.30% K	0.03%	0.13%	0.09%	0.00%	98.27%	0.00%		<u>99.05%</u> o		
TS1	171	0	0.04%	0.00%	0.81%	0.03%	0.28%	0.07%	0.01%	0.03%	0.00%	0.00%	0.01%	0.00	
DC	149	9	0.00%	- <u>8</u> 7-	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	-			-	
.I1(GID:2313)/**FLI1	132	5	0.01%	0.00%	1.44%	0.01%	0.08%	0.06%	0.00%	0.03%	0.00%	0.00%	0.06%	1	
TF4	132	1	0.00%	0.06%	0.00%	0.00%	0.02%	1.37%	7.96%	102-	0.02%	-	<u></u>	4	
GF10	132		0.01%	0.00%	-	0.01%	0.18%	0.03%	0.00%	-	0.00%		0.00%	0.01	

Literature Lab<sup>™</sup> PLUS analysis Figure 1.B.

The identity and strength of the associations as well as the underlying statistical data including the LPF (quantitative measure of co-occurrence strength) and P-value (significance based on comparison with 1000 random gene sets) are critically important in the Literature Lab<sup>TM</sup> PLUS analysis. We would like to focus on another important aspect of the platform -- the hyperlinks in the results data table that link directly to the PubMed search results involved in the analysis.

In contrast with other gene analysis platforms that function essentially as black boxes, Literature Lab<sup>TM</sup> provides direct links to the PubMed searches and literature that underlie the statistical analysis. This allows the researcher direct access to the source material in order to validate the association, through the identity of known citations, and more importantly to explore unanticipated citations uncovered by the analysis. This latter feature can be a "stepping off" point for exploring additional genes, pathways, and diseases that are not intuitive and would not otherwise be identified through simple boolean PubMed searches. In essence, Literature Lab<sup>TM</sup> can expand areas of investigation and has the potential to function as a driver of research hypotheses.

There is an established and growing body of literature surrounding GSK-3 modulation of inflammation and apoptosis in Alzheimer's and other neurological diseases. An emerging correlation between the GSK-3 pathway *Mol Psychiatry.* 2015 Mar;20(3):361-8. doi: 10.1038/mp.2014.22.



and SZ has recently been discovered, however, published work in this area is extremely limited. Literature Lab<sup>TM</sup> identified a positive association between the upregulated SZ gene set and GSK-3 and highlighted the contribution of individual genes and the underlying literature.

Literature Lab<sup>™</sup> PLUS analysis Figure 2.A.

SNCBI Resources ⊙	How To 🕑	
Pub Med.gov	PubMed	•
US National Library of Medicine National Institutes of Health		Advanced

#### Abstract -

Send to: -

Front Cell Neurosci. 2013 Mar 27;7:31. doi: 10.3389/fncel.2013.00031. eCollection 2013.

## NMDA hypofunction as a convergence point for progression and symptoms of schizophrenia.

Snyder MA<sup>1</sup>, Gao WJ.

### Author information

#### Abstract

Schizophrenia is a disabling mental illness that is now recognized as a neurodevelopmental disorder. It is likely that genetic risk factors interact with environmental perturbations to affect normal brain development and that this altered trajectory results in a combination of positive, negative, and cognitive symptoms. Although the exact pathophysiology of schizophrenia is unknown, the N-methyl-D-aspartate receptor (NMDAR), a major glutamate receptor subtype, has received great attention. Proper expression and regulation of NMDARs in the brain is critical for learning and memory processes as well as cortical plasticity and maturation. Evidence from both animal models and human studies implicates a dysfunction of NMDARs both in disease progression and symptoms of schizophrenia. Furthermore, mutations in many of the known genetic risk factors for schizophrenia suggest that NMDAR hypofunction is a convergence point for schizophrenia. In this review, we discuss how disrupted NMDAR function leads to altered neurodevelopment that may contribute to the progression and development of symptoms for schizophrenia, particularly cognitive deficits. We review the shared signaling pathways among the schizophrenia susceptibility genes DISC1, neuregulin1, and dysbindin, focusing on the AKT/GSK3β pathway, and how their mutations and interactions can lead to NMDAR dysfunction during development. Additionally, we explore what open questions remain and suggest where schizophrenia research needs to move in order to provide mechanistic insight into the cause of NMDAR dysfunction, as well as generate possible new avenues for therapeutic intervention.

KEYWORDS: NMDA receptors; gene; neurodevelopment; psychiatric disorders; schizophrenia

The positive association with the GSK-3 pathway led to linked source material such as above. Expanded searches of genes in the set that are involved in the GSK-3 pathway lead to informative source material offering a plethora of interesting associations and areas for further exploration.

Literature Lab<sup>™</sup> also identified a number of critical pathways in the downregulated gene set which have been previously associated with SZ including: Notch signaling (involved in neuron maturation), synapse formation, Ca signaling, axon guidance, and long-term depression and potentiation. The ErbB signaling pathway is moderately associated with this gene set. Insufficient ErbB signaling in humans is associated with the development of neurodegenerative diseases, such as multiple sclerosis and Alzheimer's disease, while excessive ErbB signaling is associated with the development of a wide variety of types of cancer.

The fact that Literature Lab<sup>TM</sup> PLUS analysis identified pathways present in NPCs that were previously identified in post-mortem brain tissue, and hence end point status, indicates roles not only in established SZ but in the neurodevelopmental stage of the disease. Importantly, Literature Lab<sup>TM</sup> PLUS also found emerging and unexpected pathways which may provide insight pointing toward potential novel areas of inquiry.

*Mol Psychiatry*. 2015 Mar;20(3):361-8. doi: 10.1038/mp.2014.22.

© 2016 Acumenta Biotech www.acumenta.com



We have highlighted a limited set of Literature Lab<sup>™</sup> platform features – analysis of a 500 gene microarray set focusing on pathway associations and PubMed source literature. The technology encompasses many other elements of important functionality including: A term viewer allowing interrogation of co-occurrences such as disease comorbidities; selection of up to twenty term domains for analysis including pathways, diseases, pharmaco actions, tissues, etc.; comprehensive comparison of associations between two gene sets; trends in associations in several gene sets, for instance time course data, tissue specific data, etc.; clustering analysis of gene families within inputted gene lists; and many more. A more comprehensive set of features and data analysis can be seen at <u>www.acumenta.com</u> and in the <u>application note here.</u>

The literature record within PubMed is too vast (>20 M publications to date) to permit comprehensive interrogation and identification of actionable associations. Over 4 M abstracts mention one or more human genes and modern high content genomic technologies are producing data at rates that outpace meaningful interpretation. Acumenta Biotech has created Literature Lab<sup>TM</sup>, the only literature mining-based platform that identifies statistically significant associations between gene lists and key concepts in the literature. At the basic level, Literature Lab<sup>TM</sup> can explore co-occurrences between term domains, *e.g.* diseases versus pathways. At a more rigorous level, Literature Lab<sup>TM</sup> PLUS can interrogate gene lists, including those derived via high content platforms, and score the strength of each gene set / term domain association and significance based on 1000 random gene list comparison. It respects the uniqueness of each gene set and returns consistent and unique results. Literature Lab<sup>TM</sup> identifies significant associations in a time efficient manner and reveals concepts and relationships in the literature that other gene analysis platforms cannot.

Damon Anderson, PhD Application Scientist 412-901-7785 danderson@acumenta.com The best way to see the power and benefits of Literature Lab<sup>™</sup> directly is to send us a gene list for complimentary analysis. <u>Click here</u> to register for a complimentary analysis or to be notified about upcoming Literature Lab<sup>™</sup> webinars.