# Structural variant in the RNA Binding Motif Protein, X-Linked 2 (RBMX2) gene found to be linked to bipolar disorder



P. Laine<sup>1</sup>, W. J. Rowell<sup>2</sup>, L. Paulin<sup>1</sup>, S. Kujawa<sup>2</sup>, D. Raterman<sup>3</sup>, G. Mayhew<sup>3</sup>, J. Wendt<sup>3</sup>, D. L. Burgess<sup>3</sup>, T. Partonen<sup>4</sup>, T. Paunio<sup>4,5</sup>, P. Auvinen<sup>1</sup>, J. M. Ekholm<sup>2</sup>

<sup>1)</sup> Institute of Biotechnology, University of Helsinki, Helsinki, Finland

- <sup>2)</sup> Pacific Biosciences, Menlo Park, CA, USA
- <sup>3)</sup> Roche Sequencing Solutions, Madison, WI 53719, USA
- <sup>4)</sup> National Institute for Health and Welfare (THL), Department of Health, Helsinki, Finland
- <sup>5)</sup> Department of Psychiatry, University of Helsinki

## Introduction

Here, we have used long-read Single Molecule, Real-Time (SMRT) sequencing to characterize a ~12 Mb genomic region on chromosome Xq24q27 significantly linked (lod score=3.54) to bipolar disorder in an extended family from a genetic sub-isolate.

This family segregates bipolar disorder (BD) in at least 4 generations with 24 affected individuals out of 61 total. Thus, this family portrays a highly elevated recurrence risk compared to the general population. We selected 16 individuals from the family for targeted sequencing. The selected individuals either carried the disease haplotype, were non-carriers of the disease haplotype, or served as married-in controls. We designed a SeqCap EZ probe pool (Roche) enriching for 5-9 kb fragments spanning the entire 12 Mb region that were then sequenced to screen for candidate structural variants (SVs) that could explain the increased risk for BD in this extended family.

Figure 3. Sequencing coverage plot across the targeted 12 Mb region on chromosome Xq24-q27 for all 16 family members. Each row represents the sequencing coverage for each family member separated by ID number. On the top the chromosomal location is

	2.32																						
pZ	2.32	p22.2	p22.13	p22.11	p21.2 p.	21.1 p11.4	p11.3	p11.22	p11.1	q11.2	q13.1 q13.	s q21.1	q21.2	q21.32	q22.1	q22.3	qZ3	q24	qzə	q26.1	q26.3	q27.2	q28
_												1 mb											
mb				130 mt	b			132 mb				134 mb			1	36 mb				138 mb			
		1		1	-	I		1		1		1		1		1		1		1		1	
[0 - 124]														1000	11.0								
المددما	ALL ALL	la phi	a de la		. A makes	A REAL PROPERTY.	All and the second second		and and and	و ما الح ما م	all and the	inter and at .	ALL AND DE .	بالألطي و	بالمالكين	and all	Allocale An.	الطار مدر	ور الد الله	and the film		A line balant	al a dea de
[0 - 124]			3.44													A CONTRACTOR OF A CONTRACTOR OF A CONTRACTOR OF A CONTRACTOR A CONTRACTOR A CONTRACTOR A CONTRACTOR A CONTRACT	10						
(Thet Line)			ou konosco						-		and the second second second												
[0 - 124]																							
							diala and				dinis, sistemb				ل الدائلين		Law, 88 4					and all a bala data	
(D - 124)			ALC: NO	the state					1	over the starts			MA AND			Contraction of the second			1000		10 A.U. 4	and the second	
	-					and a set of the		L. M.		In.		-		. Miles	ل الباقي						a bia linua		
(D - 124)	and the second	. Y. W	UR V H	the state				A AND A DECK	California (California)	date, billing	And the second se		And Distance		No. O Mar	and the second	and the second second	444 6446	1 Contraction		ALC: NO.		Children, I
(0 - 124)	and the second second	des. Belef	Luth	Rada .	- a stanis	a la la la la	Hadra	a di Manana da	and the state	design the state	a state and a state of the local distance of	and the section of	And and the second	and the state of the	and the last of the	mid A.m.	all had been	And a state of	Ubtellet	and the state	tonal chan	in the second second	net la
m - 1241	Industria da da	den. dies	AL ALL	Mark In	a dubanda	anti-factore as	the state and	a shirt think the	and states at \$150	A Alteria A	a different and des table	a Maria Maria	in a bill phy	and the second	ار المالاندي	in a sub	all all a line	ALCOR DUCK	City and	the state of the	and a fit of	States Marine	and the de
-	and the second	hh phi	Mumilia	A A AL AL	A Anti-May	A summer li	in more state	a hole with the	dial note	dis a filme	al des and dess	and the state of the	And the second	de staal die staal d	and the second	<b>Line de la t</b>	and the first	diaman.	din shew	-And Lake	Land Like Chall	and a share of the second	م. المالية ال
	a de la cale de	Ass. and a	the set of the local	-		and the second se		- distantion	A REAL PROPERTY.	and the second lines		Maria Maria	-	and platfalleness	and the second	and the state	a data da a		life sector	della a silla	- Martin	diama di Barda	م <u>ماد م</u> د
<b>MARKA</b>	the sale	hamu	it and a	Acres 11	A ROBALLA	and and a	and hards	- datation	and a fear	-	allan malalasi	all and all	he and stellar	and a little file	لير الما السي	and the state	-	and Astro	a dista da se	diam'r	b-Augh-h	a salture	ALC: No. A.
	and the second	Anglese	it in the	the second		and second second	in the second	a minatera	territory in al	and the bar		undia au	a	and the second	and the ful	man	abult and	and Barris	lin at at	and a state	sharel (mark	A sublemble to	mile Hearth
[0 - 124]															A.1. 287-14								
	A								- C					and a second second			100			100			

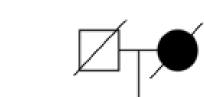
**F9** 

Altogether, 201 SVs were detected in the critically linked region. Although most of these represented common variants, three SVs emerged that showed nearperfect segregation among all bipolar disorder type I affected individuals and absence among the married-in controls. Two of these SVs were identified in genes belonging to the RNA Binding Motif Protein, X-Linked (*RBMX*) gene family - a 330 bp Alu deletion in intron 3 of the RNA Binding Motif Protein, X-Linked 2 (*RBMX2*) gene and another intergenic variant in the *RBMX* gene. The third SV was a 50 bp tandem repeat insertion in intron 1 of the Coagulation Factor IX (F9) gene. These variants are prime candidates for investigating the molecular basis of bipolar disorder in this family.

# Material: Extended family P101

Bipolar disorder I

Major Depression



displayed and at the bottom the genes and probe locations are shown. The red arrows point to the three SVs that where highlighted in this study that were located in *RBMX* and *F9* Intergenic between **RBMX2** genes. *RBMX/GPR101* genes

SV #	SV type	Size	Location	Carriers of SV	Verification methods
1.	Alu deletion	330 bp	chrX:130405824, <i>RBMX2</i> intron 3	685, 687, 688, 691, 694, 696, 698, 699, 700	PBSV workflow, PCR+Sanger, manual data inspection
2.	Tandem repeat deletion	27 bp	chrX:137012126, intergenic between genes <i>RBMX</i> and <i>GPR101</i>	685, 686, 687, 688, 691, 694, 695, 696, 698, 699, 700	PBSV workflow
3.	Tandem repeat insertion	50 <u>bp</u>	chrX: 139536239, F9 intron 1	685, 687, 688, 691, 692, 694, 695, 696, 698, 699, 700	PBSV workflow

Results

Table 1. Three SVs were highlighted in this study. These included an Alu deletion (1, red), a tandem repeat deletion (2, blue) and a tandem repeat insertion (3, black) of various sizes. Two SVs (1 and 2) were located in or in close vicinity of different RBMX gene family members, while one was located in the F9 gene. The pedigree in Figure 1 shows the segregation by number of each of the SVs.

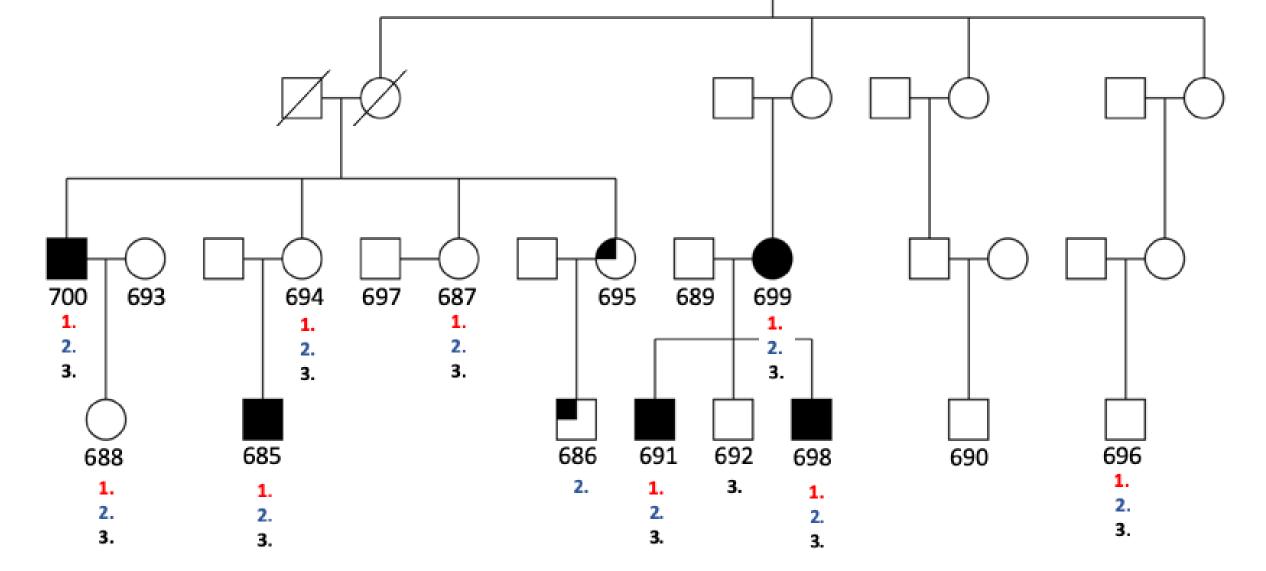


Figure 1. Family P101. Sixteen family members were included in the study. Five were affected males (bipolar disorder I: 700, 685, 691, 698, major depression: 686) and two were affected females (bipolar disorder I: 699, major depression: 695). In addition, three married-in unaffected controls were included (693, 697, 689) as well as six unaffected family members (688, 694, 687, 692, 690, 696). See **Table 1**. for explanation on the colored numbers 1-3 seen here.

# Methods: SeqCap<sup>®</sup> EZ probe-based capture combined with long-read SMRT Sequencing

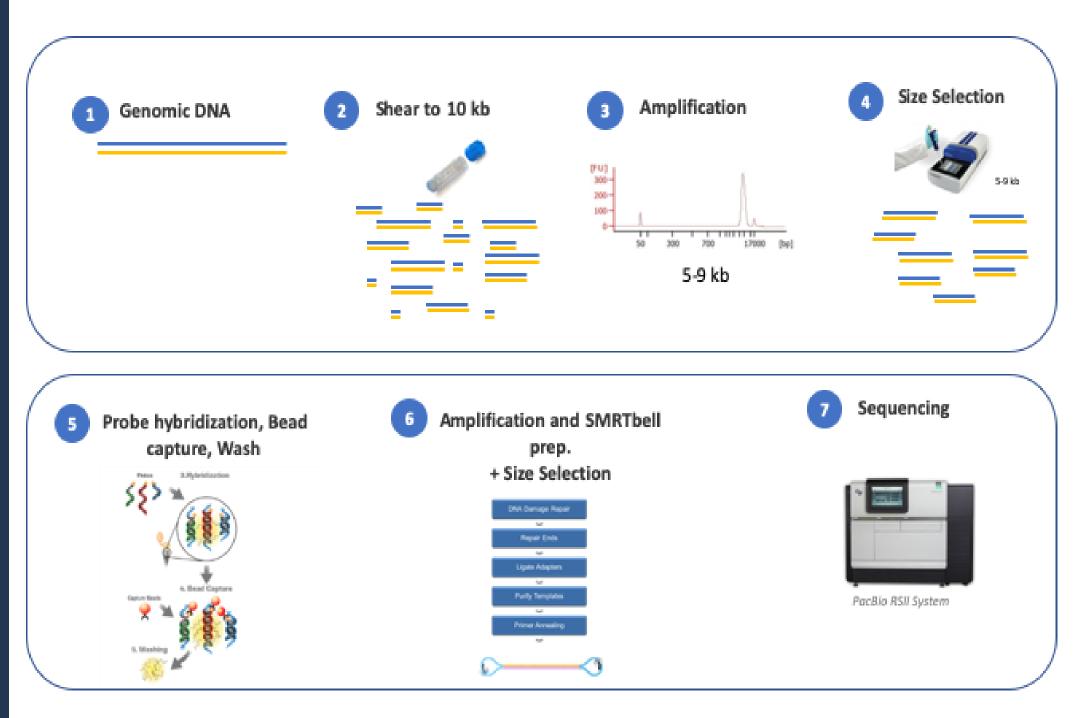


Figure 2. Workflow overview. gDNA<sup>1)</sup> was sheared<sup>2)</sup> 10 kb fragments which was followed by ligation <sup>3)</sup> of adapters,

amplification and size selection <sup>4)</sup> using the BluePippin System. Next, the SeqCap<sup>®</sup> EZ capture <sup>5)</sup> probes where hybridized, followed by the bead capture step and a washing step to remove any nonspecific and unbound molecules. Finally, the sequencing library was generated <sup>6)</sup> which was size selected with BluePippin before sequencing<sup>7</sup>) on the PacBio RS II system.

#### Conclusions

#### Summary

Long-read SMRT Sequencing was applied to a 12 Mb genomic region on chromosome Xq24-q27 linked to BD in an extended family from an isolated population

- 16 key individuals from the extended family were selected for the study
- Roche's hybridization-based target enrichment of 5-9 kb fragments spanning the region was used in combination with PacBio long-read sequencing
- Altogether, 201 SVs were detected most representing common variants, however 3 SVs emerged that showed near-perfect segregation among all bipolar disorder type I affected individuals

## Conclusions

There is additive evidence that the 330bp *Alu* deletion in intron 3 of the *RBMX2* gene may be involved in the disease development of BD type I in this extended P101 family

- Near-perfect segregation with BD type 1 in family P101
- *RBMX2* play a central role in brain development and function and has previously been associated with mental retardation and epilepsy
- Deleterious Alu activity is associated with at least 37 neurological and neurodegenerative disorders

Further studies are warranted to fully understand whether the RBMX2 gene pathway is involved in the development of BD type I and to confirm the functionality of the identified variant

Acknowledgements

The authors would like to thank all study participants for their generous participation at THL Biobank, Helsinki, Finland

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2019 by Pacific Biosciences logo, PacBio, SMRT, SMRTbell, Iso-Seq, and Sequel are trademarks of Pacific Biosciences. BluePippin and SageELF are trademarks of Sage Science. NGS-go and NGSengine are trademarks of Agilent Technologies Inc. All other trademarks are the sole property of their respective owners.