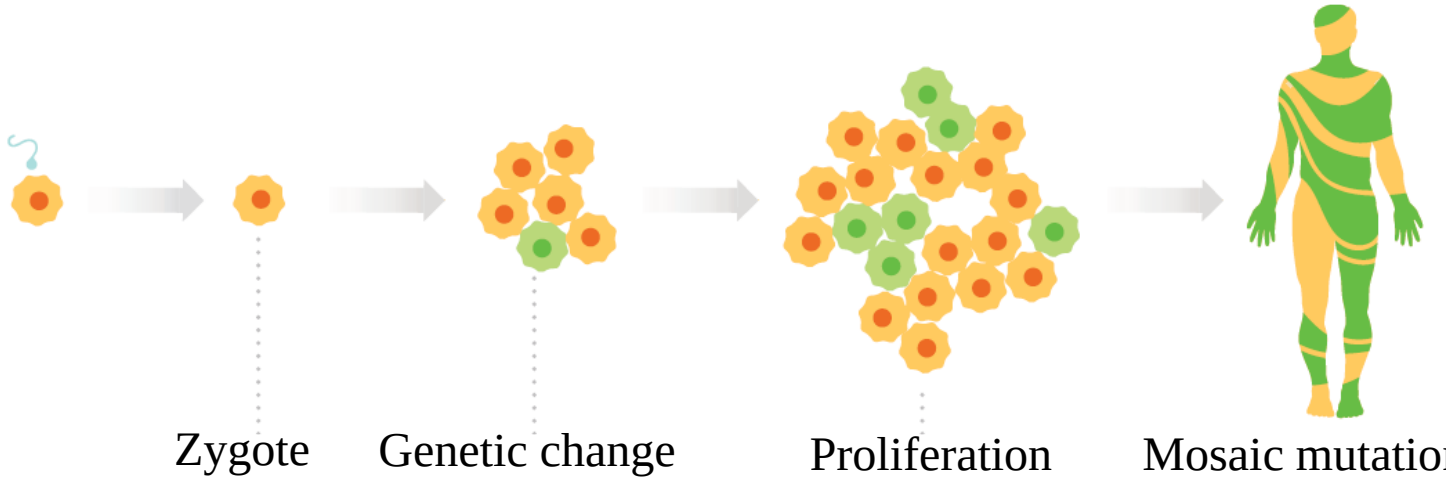
Deep amplicon resequencing identified parental mosaicism for approximately 10% "*de novo*" *SCN1A* mutations in Dravet Syndrome families and was capable of multiple validations of mosaicism

Xiaoxu Yang, Ph. D. candidate

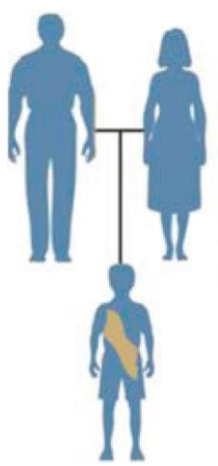
Dr. Liping Wei's Laboratory
Center for Bioinformatics, School of Life Sciences
Peking University, Beijing, China

Oct 18, 2017

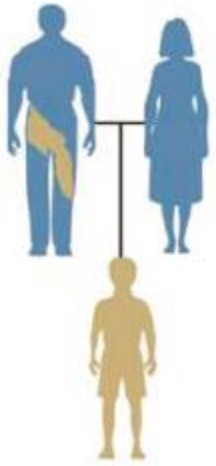
Mosaicism contributes to the **etiology** of diseases and the **origin** of mutations

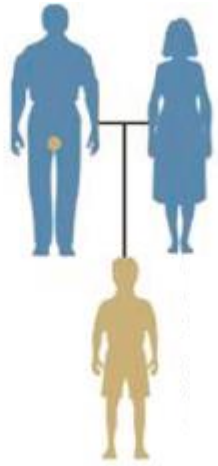


(Servick, *Science*, 2014)



Cases of deleterious mosaicism reported in more than 50 disorders





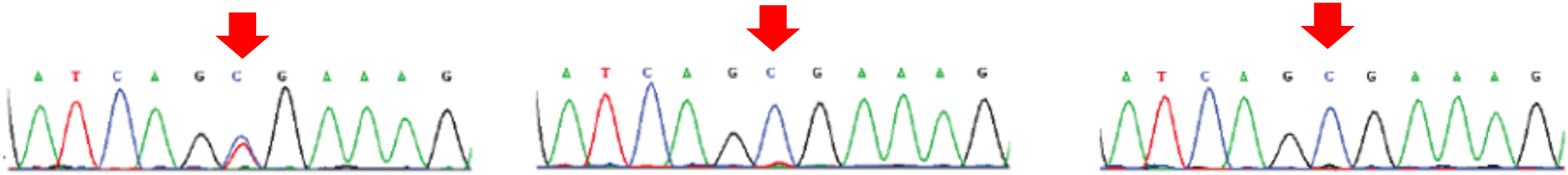
Cases of parental mosaicism reported in more than 100 disorders

(Freed, Stevens and Pevsner, *Genes*, 2014)

The limitation of Sanger sequencing might lead clinicians to mistake potential postzygotic single-nucleotide mosaicisms (**pSNMs**) for "*de novo*" mutations

Sanger sequencing

***De novo*?**



Proband

Father

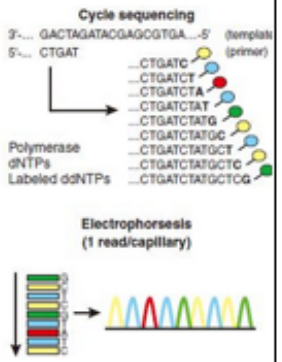
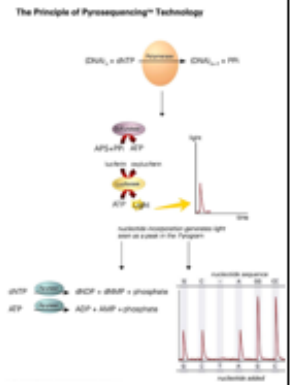
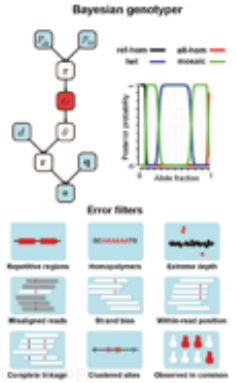
Mother

How to screen and quantify mosaicisms in non-cancer disorders?

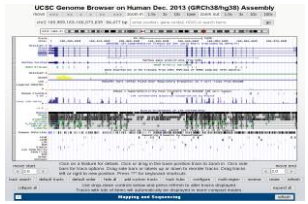
How many of the "*de novo*" mutations were undetected parental mosaicisms?

(Modified from Xu, *et al.*, *Hum Mutat*, 2015)

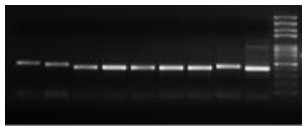
Technologies to identify and validate mosaicism

	Sanger sequencing	Pyrosequencing	Micro-droplet digital PCR	PASM	MosaicHunter
Schmatic diagram					
Features	Coventional method widely used	Former gold standard for mosaicism	Ultra-sensitive with Single molecule definition	The first semiconductor sequencing based cost effective ethod. Cost effective of large-scale screen for pSNMs.	A powerful Bayesian based bioinformatic tool for WES and WGS data in unpaired, paired and trio sample
Detection limit	10%	5%	0.01%	0.5%	Could be adjusted with imput
Publication	(Sanger, *et al. Proc. Natl. Acad. Sci. USA*, 1977)	(Nyren, Petersson, and Uhlent., *Anal. Biochem.*,1993)	(Zhong, *et al*, *Lab on a Chip*, 2011)	(Xu, Yang, and Wu,*et al. Hum. Mutat.*,2015)	(Huang *et al., Cell Res.* 2014 and Huang *et al.Nucleic, Acids, Res.* 2017)

PGM Amplicon Sequencing of Mosaicism (**PASM**) for detection and quantification of the mutant allelic fraction (**MAF**) of mosaicism

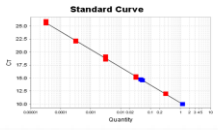


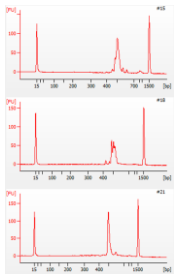
Target region



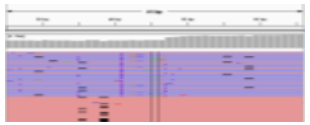
(Nested) **PCR capture**

Quality controls





Library construction, Template preparations and PGM semiconductor sequencing

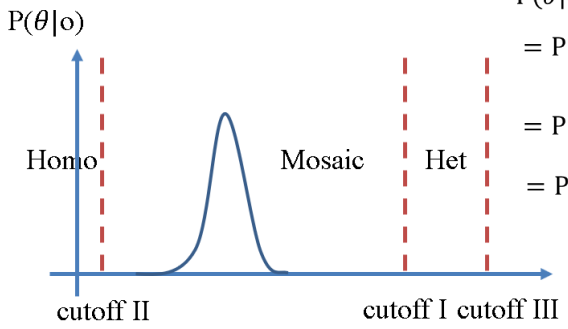


Sequencing data

https://www.thermofisher.com/cn/en/home/life-science/sequencing/next-generation-sequencing.html

Filters:
Base quality filter
Read depth filter
Strand bias filter

Bayesian model simplified from MosaicHunter



$$P(\theta|o) \propto P(\theta)P(o|\theta)$$
$$= P(\theta)\sum_r P(o,r|\theta)$$
$$= P(\theta)\sum_r P(r|\theta)\,P(o|r)$$
$$= P(\theta)\sum_r P(r|\theta;n)\,P(o|r;q)$$

θ: theoretical MAF
o: mutant reads
n: total reads
r: "real" allele count
$P(\theta)$: prior
$P(r|\theta;n)$: likelihood
$P(o|r;q)$: summarized p

Scripts and packages for calculation

MAP estimation for mutant allelic fractions (**MAF**s)

95% Credible Intervals (**CI**s)

95% CI of the minor allele under cutoff I and above cutoff II , or above cutoff III and under cutoff IV

Yes

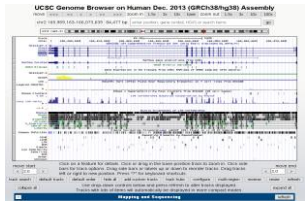
Mosaic

No

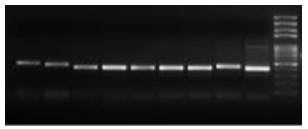
Homo/heterozygous

(Modified from Xu, *et al.*, *Hum Mutat*, 2015)

PGM Amplicon Sequencing of Mosaicism (**PASM**) for detection and quantification of the mutant allelic fraction (**MAF**) of mosaicism

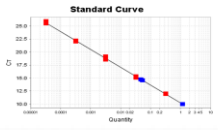


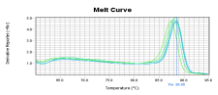
Target region

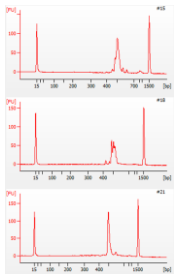


(Nested) **PCR capture**

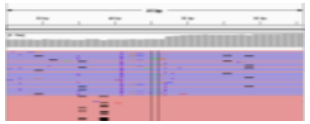
Quality controls







Library construction, Template preparations and PGM semiconductor sequencing



Sequencing data

https://www.thermofisher.com/cn/en/home/life-science/sequencing/next-generation-sequencing.html

Filters:
Base quality filter
Read depth filter
Strand bias filter

Bayesian model simplified from MosaicHunter

$$P(\theta|o) \propto P(\theta)P(o|\theta)$$
$$= P(\theta)\sum_r P(o,r|\theta)$$
$$= P(\theta)\sum_r P(r|\theta)\,P(o|r)$$
$$= P(\theta)\sum_r P(r|\theta;n)\,P(o|r;q)$$

θ: theoretical MAF
o: mutant reads
n: total reads
r: "real" allele count
$P(\theta)$: prior
$P(r|\theta;n)$: likelihood
$P(o|r;q)$: summarized p

Scripts and packages for calculation

MAP estimation for mutant allelic fractions (**MAF**s)

95% Credible Intervals (**CI**s)

95% CI of the minor allele under cutoff I and above cutoff II , or above cutoff III and under cutoff IV

Yes

Mosaic

No

Homo/heterozygous

(Modified from Xu, *et al.*, *Hum Mutat*, 2015)

PGM Amplicon Sequencing of Mosaicism (**PASM**) for detection and quantification of the mutant allelic fraction (**MAF**) of mosaicism

Target region

(Nested) **PCR capture**

Quality controls

Library construction, Template preparations and PGM semiconductor sequencing

Sequencing data

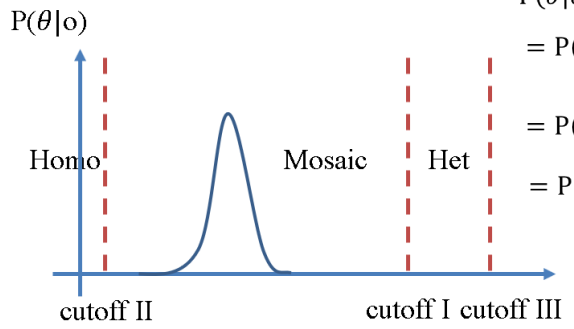
https://www.thermofisher.com/cn/en/home/life-science/sequencing/next-generation-sequencing.html

Filters:
Base quality filter
Read depth filter
Strand bias filter

Bayesian model simplified from MosaicHunter

θ: theoretical MAF
o: mutant reads
n: total reads
r: "real" allele count
$P(\theta)$: prior
$P(r|\theta;n)$: likelihood
$P(o|r;q)$: summarized p

$$P(\theta|o) \propto P(\theta)P(o|\theta)$$
$$= P(\theta)\sum_r P(o,r|\theta)$$
$$= P(\theta)\sum_r P(r|\theta)\,P(o|r)$$
$$= P(\theta)\sum_r P(r|\theta;n)\,P(o|r;q)$$



Scripts and packages for calculation

MAP estimation for mutant allelic fractions (**MAF**s)

95% Credible Intervals (**CI**s)

95% CI of the minor allele under cutoff I and above cutoff II , or above cutoff III and under cutoff IV

Yes

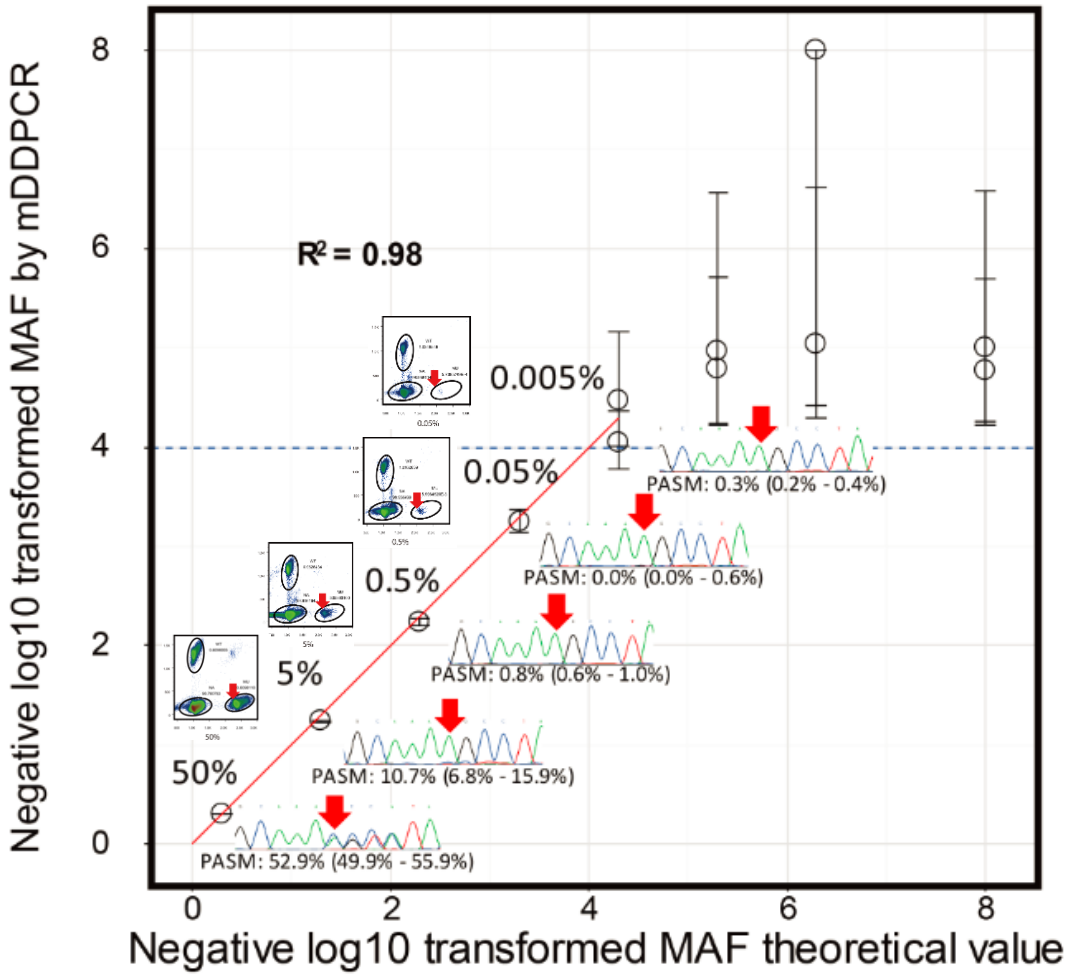
Mosaic

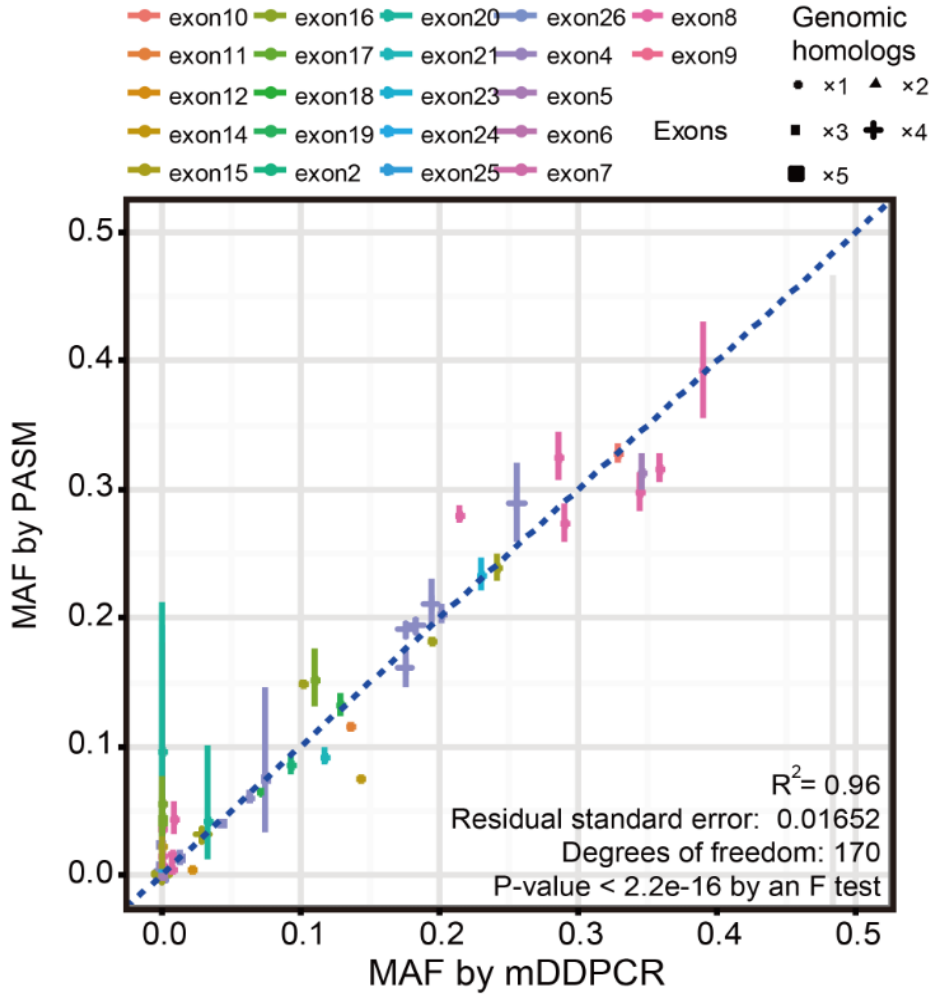
No

Homo/heterozygous

(Modified from Xu, *et al.*, *Hum Mutat*, 2015)

MAFs measured by PASM showed strong accordance with results form corrected micro-droplet digital PCR (**mDDPCR**)



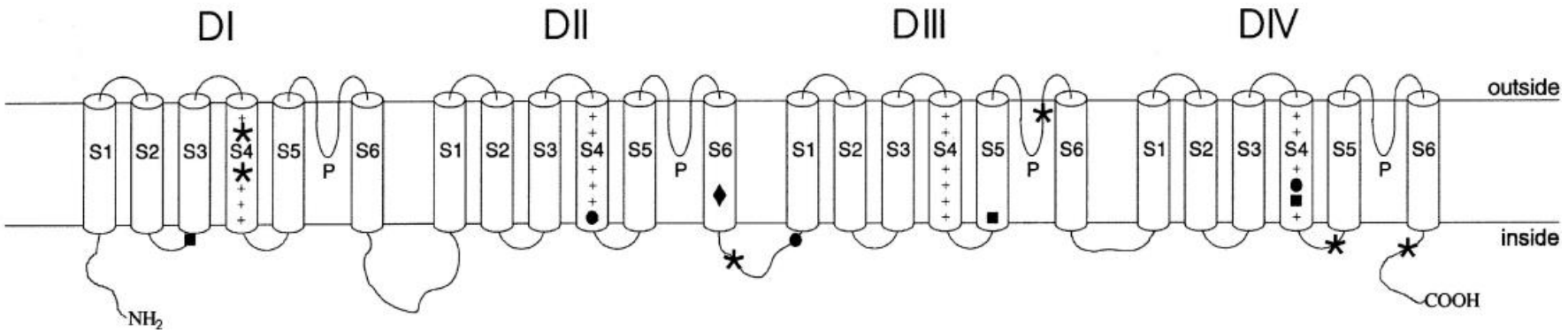


(under review)

Dravet syndrome (**DS**) caused by *SCN1A* mutations

- Characteristics of Dravet Syndrome (MIM#607208, Severe Myoclonic Epilepsy of Infancy)
  - Seizure onset within 1 year of age; multiple seizure types after 1 year of age
  - Seizures are commonly fever-sensitive and refractory to drug therapy
  - Normal early development; psychomotor development delay after seizure onset

Sodium Channel Neuronal type I alpha subunit (*SCN1A*)



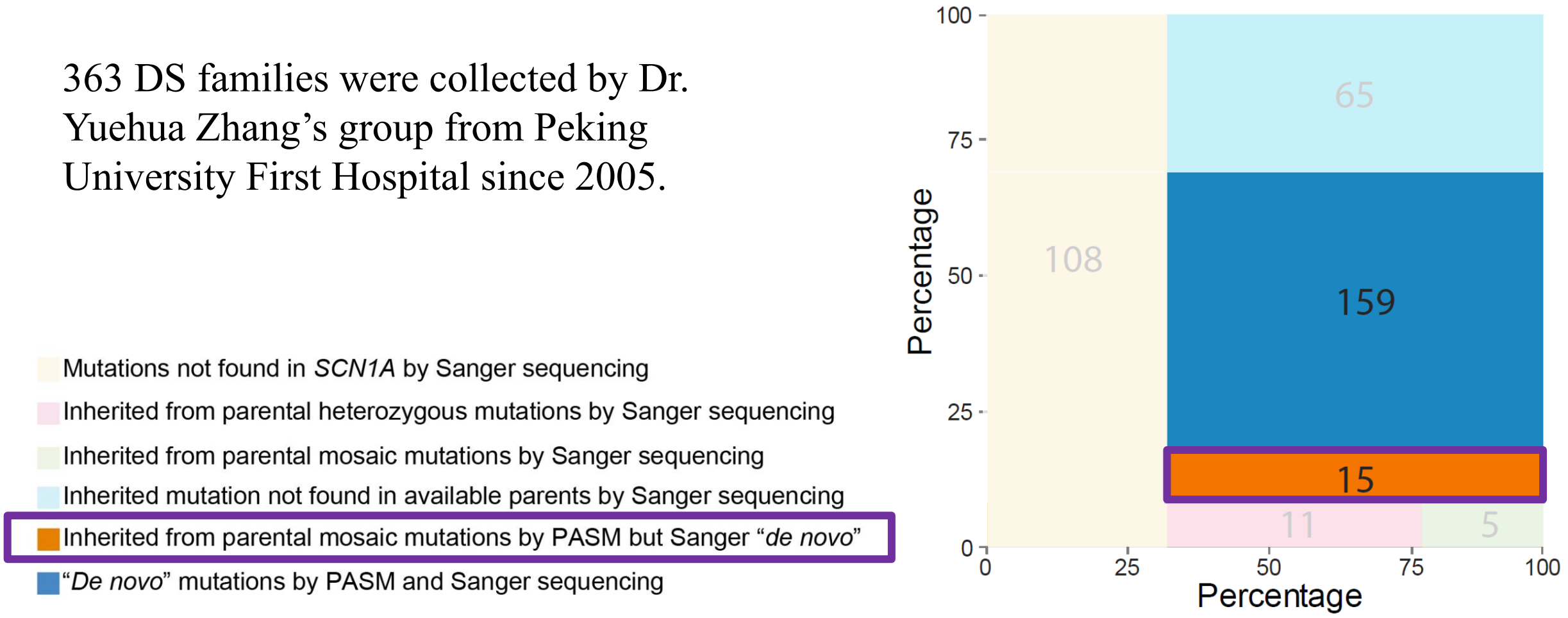
(Claes *et al*., *Am J Hum Genet*, 2001)

- Genetics of DS :
  - **More than 70%** DS probands have rare missense/Loss-of-Function mutations on ***SCN1A***
  - **90-95%** of the mutations are ***"de novo"*** by Sanger sequencing
  - Over 10% family members of the proband had milder phenotype

(Escayg *et al*., *Nat Genet*, 2000 ; Dravet, *Epilepsia*, 2011)

PASM identified **15 parental mosaicisms** in 174 Sanger sequencing regarded "*de novo*" DS families

363 DS families were collected by Dr. Yuehua Zhang's group from Peking University First Hospital since 2005.



(Xu, *et al.*, *Hum Mutat*, 2015)

List of the 75% (15 of 20) parental mosaicisms detected by PASM but missed by Sanger sequencing

Family	Proband mutation information				Mosaic parent information			Mosaic Related Phenotype	Mosaic site information				
	Chromosome	Position [a]	Nucleotide Variations [b]	Amino Acid Variation [b]	Parent of Origin	Reference Allele	Alternative Allele	Epileptic symptoms in parents	MAFs by PASM	95% Credible Interval		Validation	
										Lower bound	Upper bound	Pyrosequencing	mDDPCR
DS017	chr2	166848438	c.5347G>A	A1783T	father	C	T	Father, FS before 5	4.0%	3.8%	4.1%	12%	4.41%
DS027	chr2	166915126	c.337C>A	P113T	father	G	T	Father, several FS at the early age	25.3%	22.3%	28.5%	43%	-
DS035	chr2	166894440	c.2792G>A	R931H	father	C	T	Neither	15.0%	14.8%	15.2%	16%	10.24%
DS094	chr2	166848852	c.4933C>T	R1645*	father	G	A	Neither	1.3%	0.8%	1.9%	3%	1.42%
DS101	chr2	166848230	c.5555T>C	M1852T	father	A	G	Neither	6.1%	5.6%	6.7%	26%	6.31%
DS104	chr2	166904137	c.1170+1G>T	-	mother	G	T	Neither	1.1%	0.9%	1.4%	6%	-
DS117	chr2	166895930	c.2589+3A>T	-	mother	T	A	Neither	2.3%	2.0%	2.5%	-	-
DS125	chr2	166868765	c.3733C>T	R1245*	father	G	A	Neither	6.6%	6.2%	6.9%	12%	7.15%
DS128	chr2	166868765	c.3733C>T	R1245*	mother	G	A	Neither	13.2%	12.4%	14.1%	19%	13.02%
DS130	chr2	166868772	c.3726_3727insAT	D1243fsX1270	father	A	T	Neither	3.3%	2.8%	3.9%	-	-
DS136	chr2	166859043	c.4223G>A	W1408*	mother	C	T	Mother, undefined epilepsy	9.2%	8.5%	9.9%	22%	11.71%
DS164	chr2	166915194	c.269T>C	F90S	father	A	G	Father, FS at the early age	8.6%	7.9%	9.4%	15%	9.32%
DS166	chr2	166894396	c.2836C>T	R946C	father	G	A	Neither	3.1%	3.1%	3.2%	6%	3.28%
DS188	chr2	166894554	c.2678T>A	L893*	mother	A	T	Neither	6.3%	1.2%	16.3%	23%	-
DS206	chr2	166901776	c.1439_1442delCAGA	S481fs*488	father	G	A	Neither	10.7%	9.3%	12.3%	-	-

[a] Position coordinates were based on the UCSC human reference genome version hg19.

[b] Nucleotide and amino acid variations were based on RefSeq sequence NM_001165963.1.

(Xu, *et al.*, *Hum Mutat*, 2015)

List of the 75% (15 of 20) parental mosaicisms detected by PASM but missed by Sanger sequencing

Family	Proband mutation information				Mosaic parent information			Mosaic Related Phenotype	Mosaic site information				
	Chromosome	Position [a]	Nucleotide Variations [b]	Amino Acid Variation [b]	Parent of Origin	Reference Allele	Alternative Allele	Epileptic symptoms in parents	MAFs by PASM	95% Credible Interval		Validation	
										Lower bound	Upper bound	Pyrosequencing	mDDPCR
DS017	chr2	166848438	c.5347G>A	A1783T	father	C	T	Father, FS before 5	4.0%	3.8%	4.1%	12%	4.41%
DS027	chr2	166915126	c.337C>A	P113T	father	G	T	Father, several FS at the early age	25.3%	22.3%	28.5%	43%	-
DS035	chr2	166894440	c.2792G>A	R931H	father	C	T	Neither	15.0%	14.8%	15.2%	16%	10.24%
DS094	chr2	166848852	c.4933C>T	R1645*	father	G	A	Neither	1.3%	0.8%	1.9%	3%	1.42%
DS101	chr2	166848230	c.5555T>C	M1852T	father	A	G	Neither	6.1%	5.6%	6.7%	26%	6.31%
DS104	chr2	166904137	c.1170+1G>T	-	mother	G	T	Neither	1.1%	0.9%	1.4%	6%	-
DS117	chr2	166895930	c.2589+3A>T	-	mother	T	A	Neither	2.3%	2.0%	2.5%	-	-
DS125	chr2	166868765	c.3733C>T	R1245*	father	G	A	Neither	6.6%	6.2%	6.9%	12%	7.15%
DS128	chr2	166868765	c.3733C>T	R1245*	mother	G	A	Neither	13.2%	12.4%	14.1%	19%	13.02%
DS130	chr2	166868772	c.3726_3727insAT	D1243fsX1270	father	A	T	Neither	3.3%	2.8%	3.9%	-	-
DS136	chr2	166859043	c.4223G>A	W1408*	mother	C	T	Mother, undefined epilepsy	9.2%	8.5%	9.9%	22%	11.71%
DS164	chr2	166915194	c.269T>C	F90S	father	A	G	Father, FS at the early age	8.6%	7.9%	9.4%	15%	9.32%
DS166	chr2	166894396	c.2836C>T	R946C	father	G	A	Neither	3.1%	3.1%	3.2%	6%	3.28%
DS188	chr2	166894554	c.2678T>A	L893*	mother	A	T	Neither	6.3%	1.2%	16.3%	23%	-
DS206	chr2	166901776	c.1439_1442delCAGA	S481fs*488	father	G	A	Neither	10.7%	9.3%	12.3%	-	-

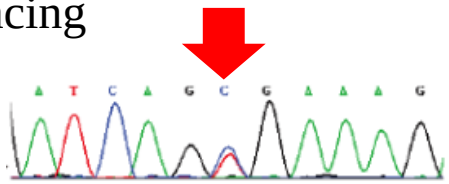
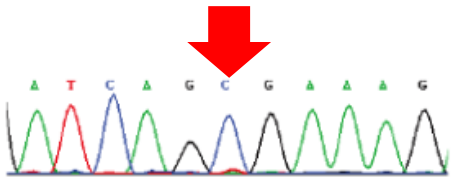
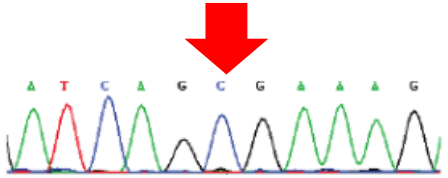
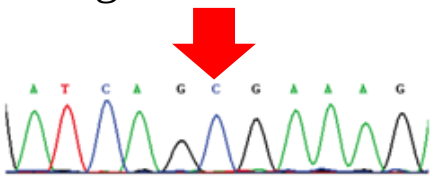
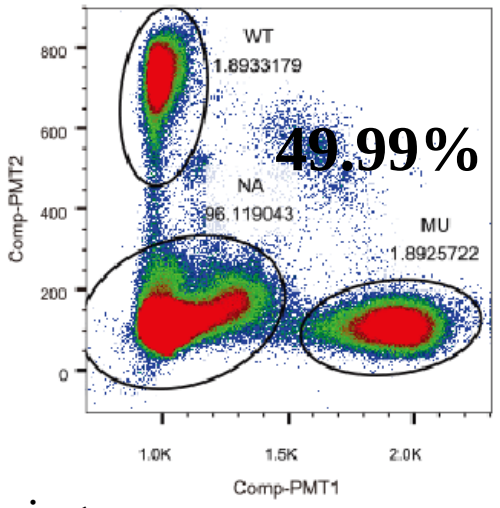
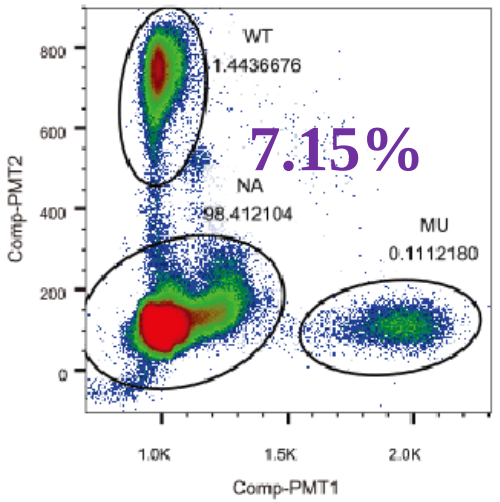
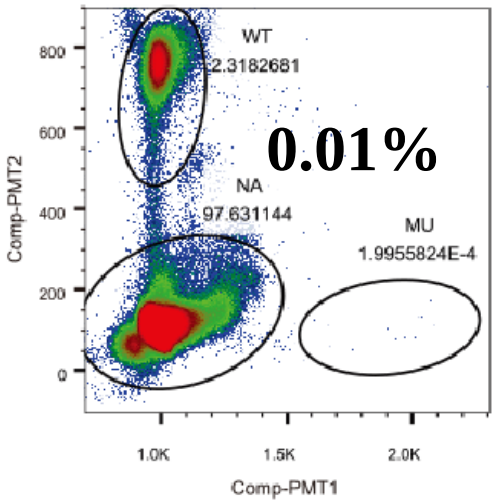
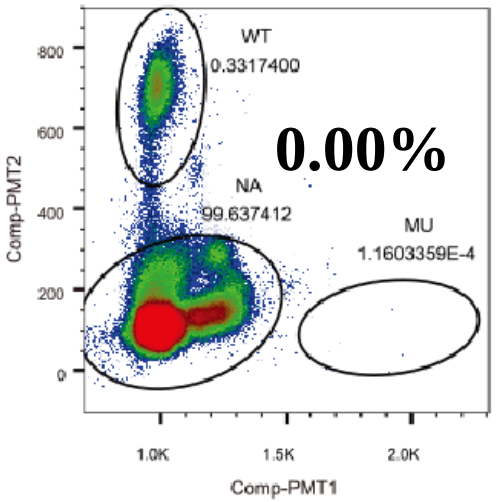
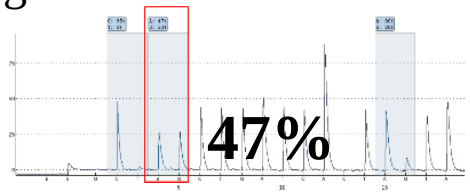
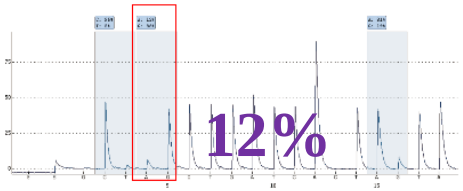
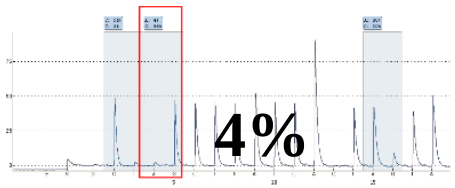
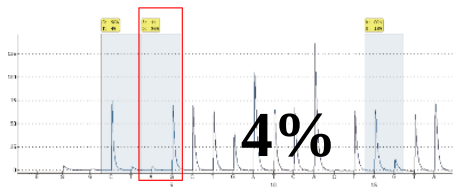
[a] Position coordinates were based on the UCSC human reference genome version hg19.

[b] Nucleotide and amino acid variations were based on RefSeq sequence NM_001165963.1.

(Xu, *et al.*, *Hum Mutat*, 2015)

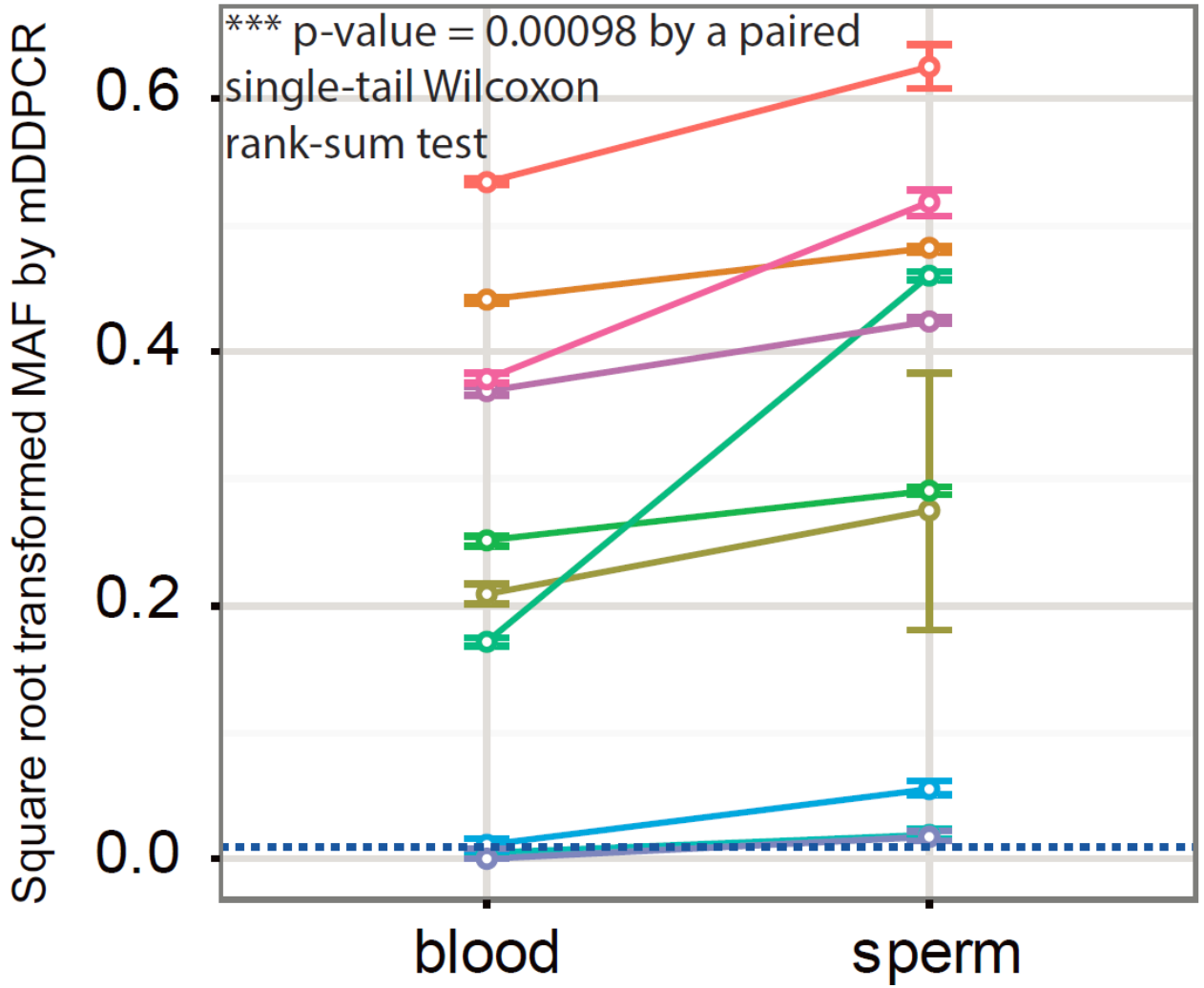
Validated pSNMs detected by PASM and missed by Sanger sequencing

6.6% by PASM

	Proband	Father	Mother	Negative control
Sanger sequencing				
mDDPCR				
Pyrosquencing				

(Modified from Xu, *et al.*, *Hum Mutat*, 2015)

mDDPCR analysis of 56 paternal sperm sample found the MAFs are significantly **higher** in paternal sperm than blood in DS families

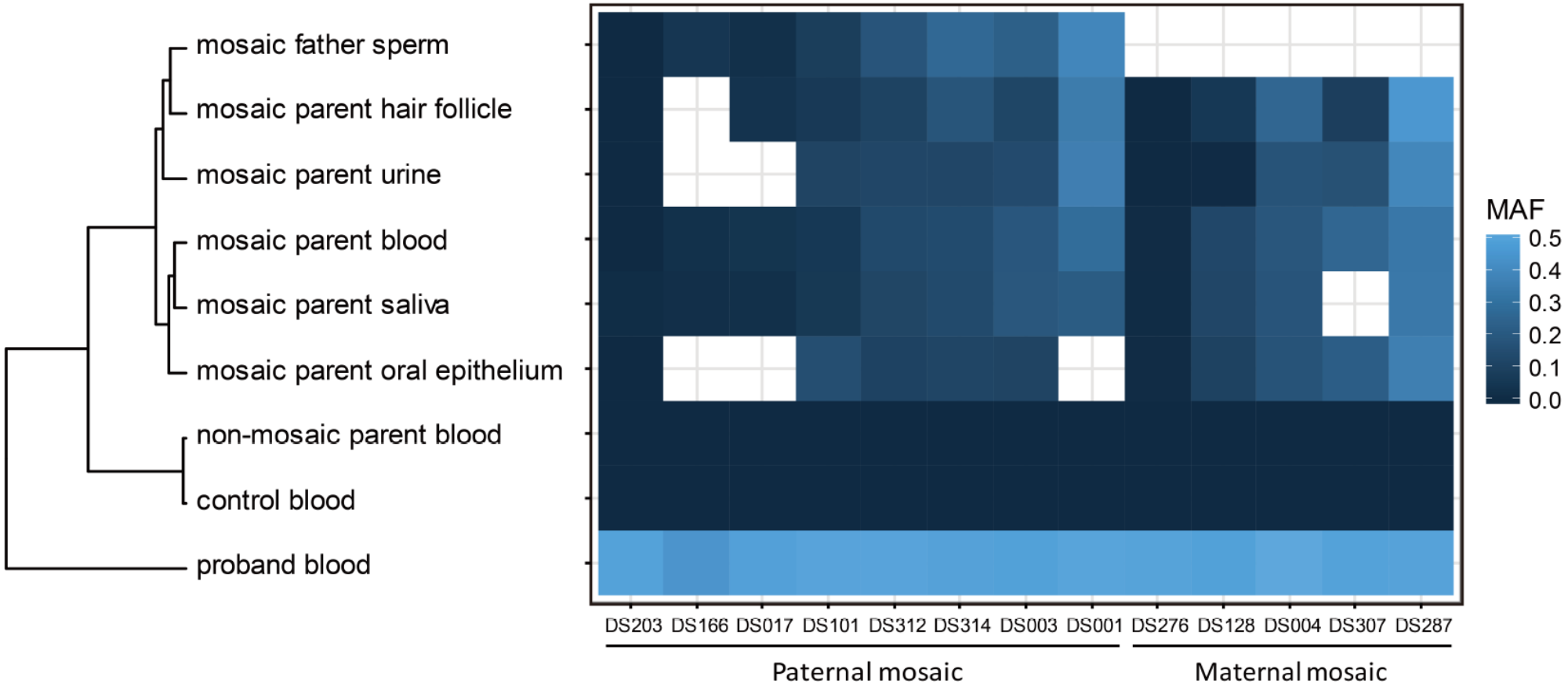


In three cases, mosaicism were detected by mDDPCR with ultra-low fraction and only detectable in sperm but not in blood.

After conditional probability correction, **P' = 0.033**

(under review)

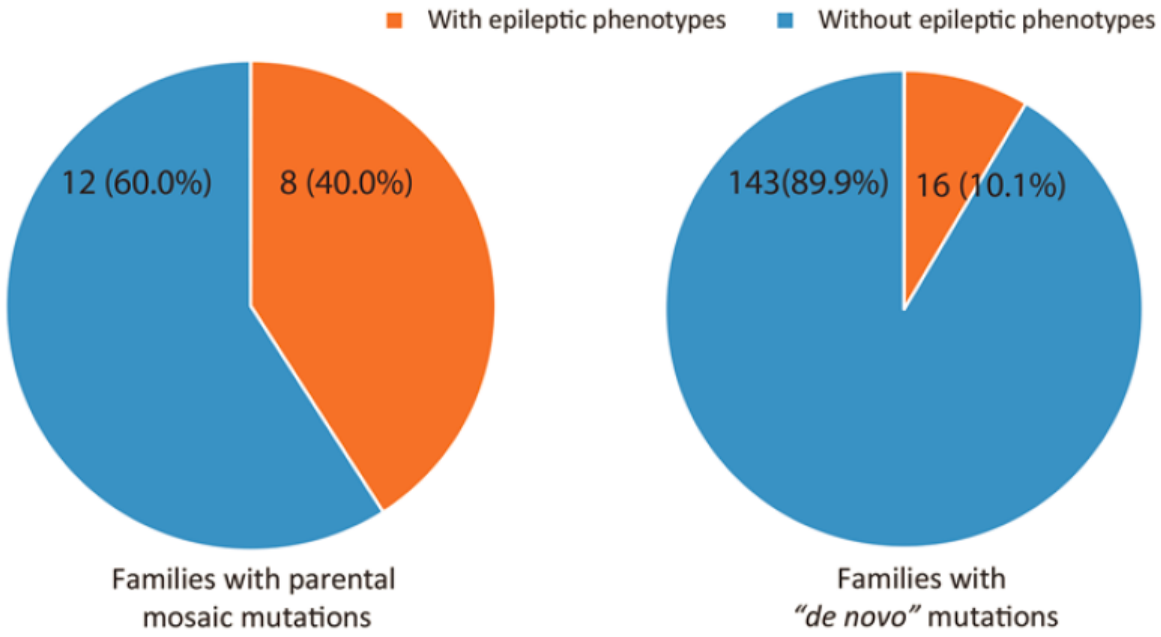
Parental multiple peripheral samples were collected in **75%** fathers paternal **sperm** were measured with the **highest** MAF



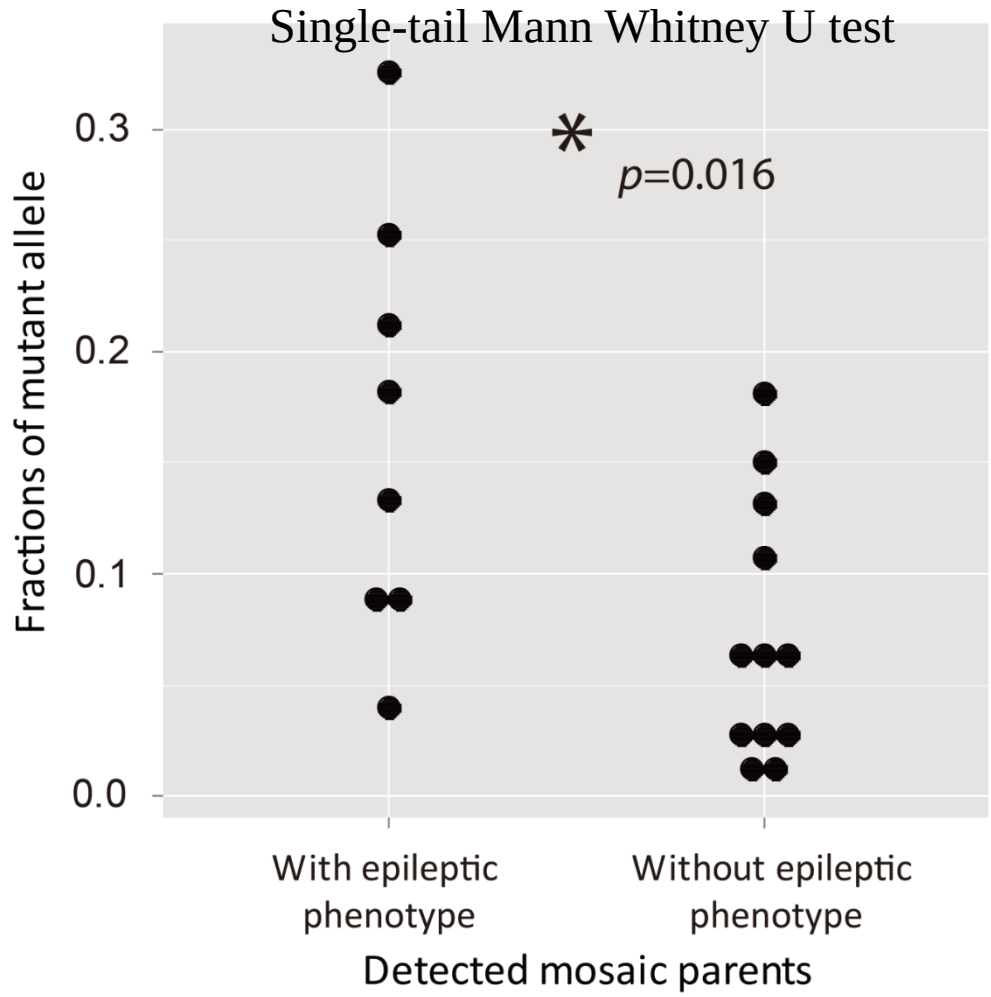
97.78% multiple parental tissues were detected with mosaicism.

(under review)

DS mosaic parents with milder epileptic phenotypes have significant higher MAF

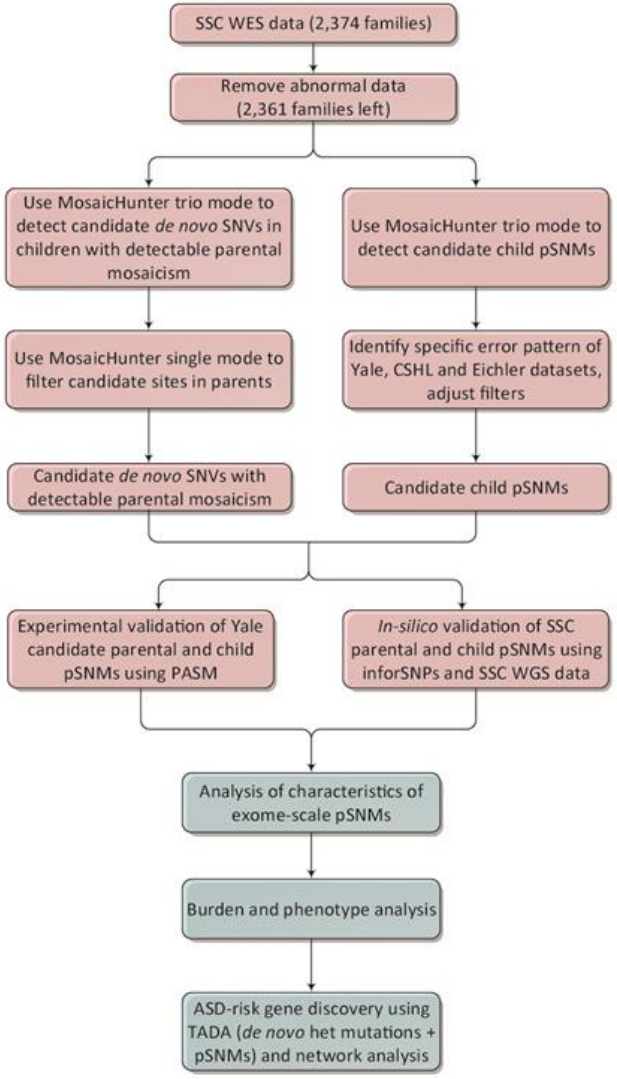


* P = 0.0143
Odds Ratio = 5.8657
By a Fisher's Exact Test

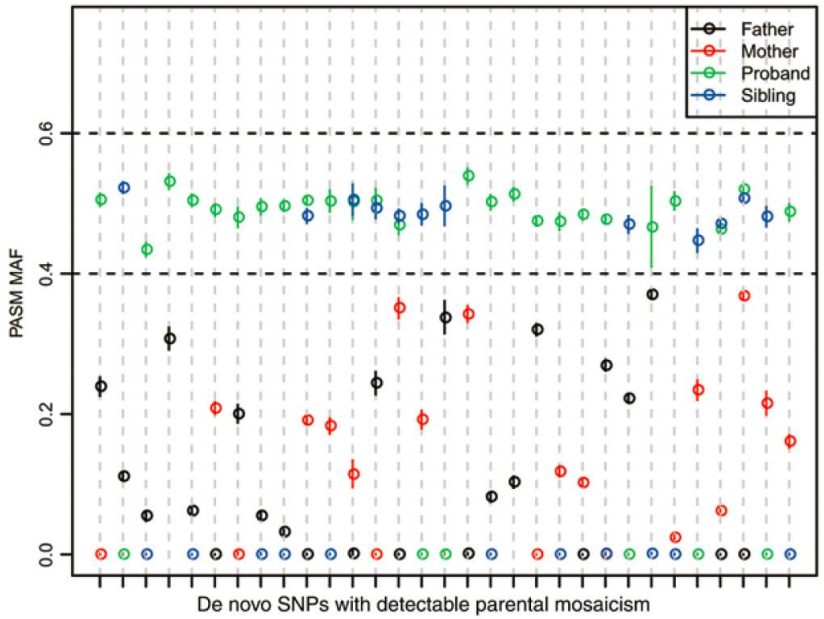


(Xu, *et al.*, *Hum Mutat*, 2015)

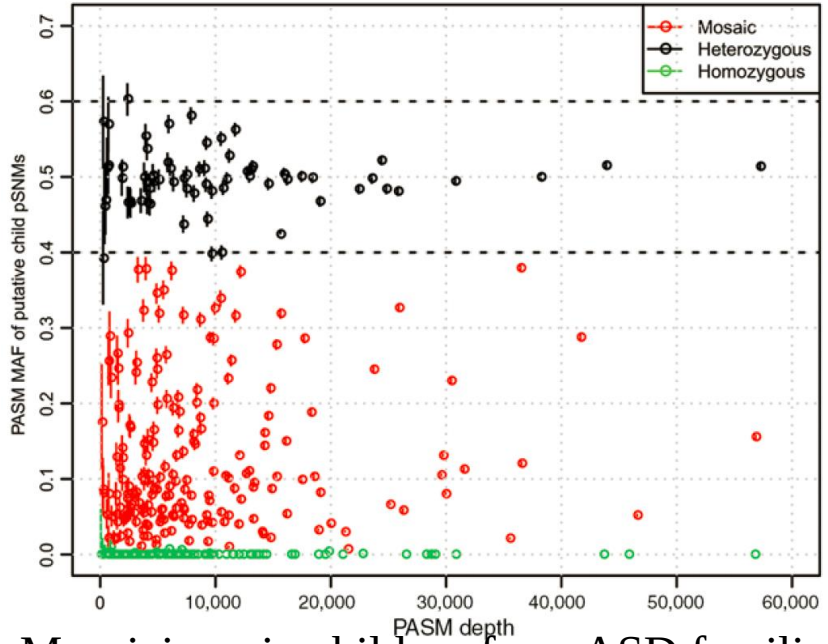
PASM validated 199/376 child mosaicisms and 31/71 parental mosaicisms from 753 blood-DNA-available families with autism spectrum disorders (**ASD**)



MosaicHunter, a bioinformatic software we developed for pSNMs was used to detect WES data of 2361 ASD families from the Simons Simplex Collection



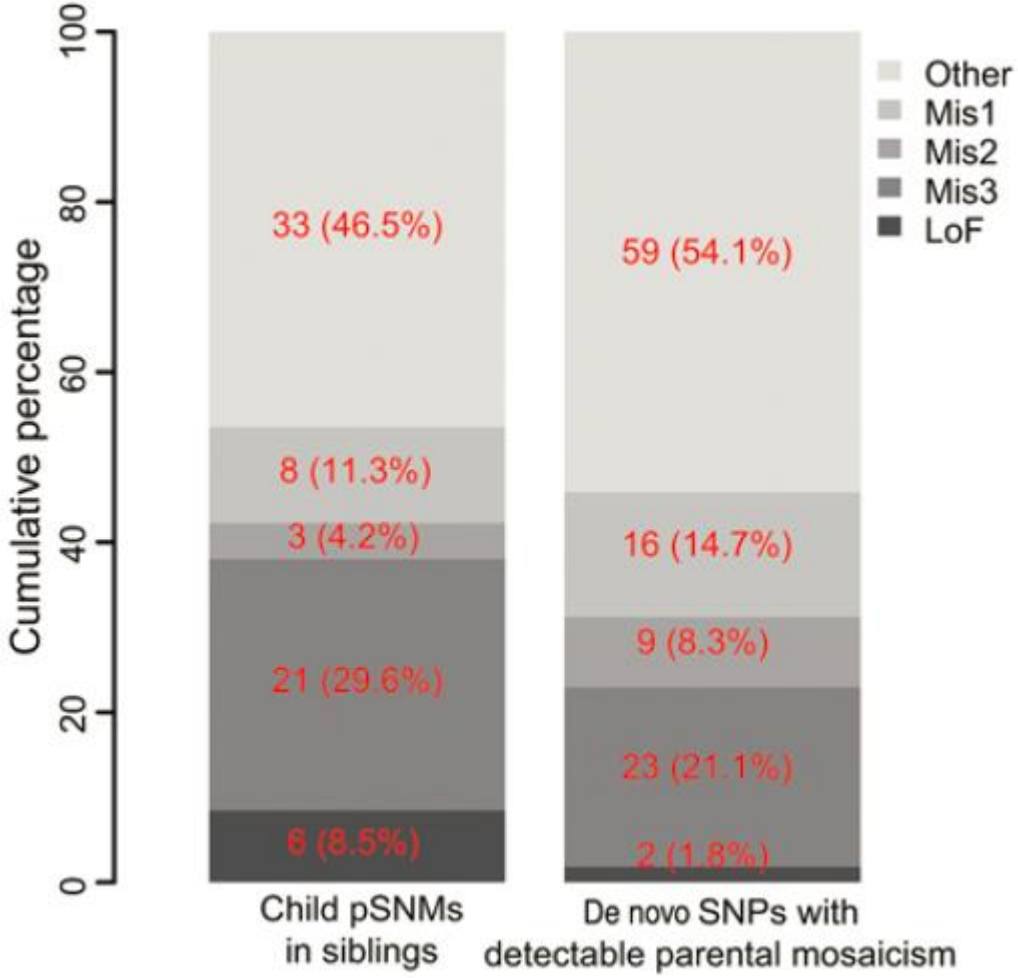
Parental mosaicism from ASD families

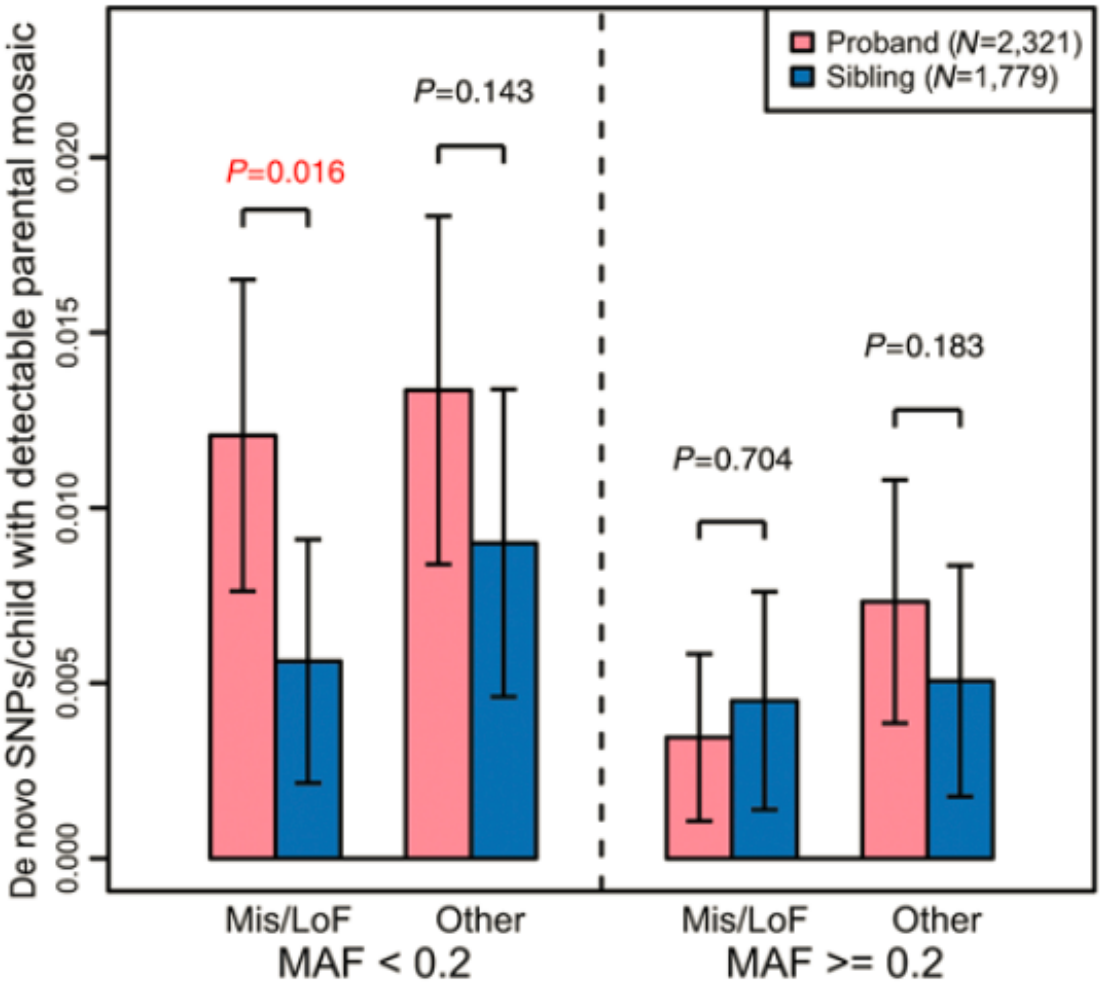


Mosaicisms in children from ASD families

(Modified from Dou, *et al.*, *Hum Mutat*, 2017)

De novo SNVs with detectable parental mosaicism with low MAFs showed two fold enrichment in ASD probands versus unaffected siblings





(Dou, *et al.*, *Hum Mutat*, 2017)

Conclusions

- Parental mosaicism is responsible for a considerable proportion in seemingly "*de novo*" mutations in Mendelian disorders such as DS
- Mosaicism increased the risk of ASD by approximately 6%
- The framework containing MosaicHunter, PASM and mDDPCR we built brought new insights into the origin, transmission, and effect of mutations

Acknowledgements

Dr. Liping Wei
weilp@mail.cbi.pku.edu.cn

Dr. Yuehua Zhang
zhangyhdr@126.com

Center for bioinformatics, School of Life Sciences, Peking University
Mr. Adam Yongxin Ye
Dr. August Yue Huang
Dr. Yanmei Dou
Mr. Ziyi Li

Human Genetic Resources Core Facility, Center for Life Sciences
Dr. Qixi Wu
Mrs. Zhe Yu
Mr. Sheng Wang

Department of pediatrics, Peking University First Hospital
Dr. Xiaojing Xu
Dr. Aijie Liu
Dr. Zhichao Zhang
Dr. Xiru Wu

Beijing Institute of Genomics, Chinese Academy of Sciences
Dr. Kaile Wang

中华人民共和国科学技术部
Ministry of Science and Technology of the People´s Republic of China

国家自然科学基金委员会

北京市科学技术委员会
Beijing Municipal Science & Technology Commission

Deep amplicon resequencing identified parental mosaicism for approximately 10% "*de novo*" *SCN1A* mutations in Dravet Syndrome families and was capable of multiple validations of mosaicism

Xiaoxu Yang, Ph. D. candidate

Dr. Liping Wei's Laboratory
Center for Bioinformatics, School of Life Sciences
Peking University, Beijing, China

Oct 18, 2017

Ethics Statement

- This study and its sample collection and detection procedures were approved by the Institutional Review Board at Peking University (IRBPU) and the Ethics Committee of Peking University First Hospital.

- Project number: IRB00001052-11087

- Written informed consent was provided by participants or their statutory guardians before enrollment.

Appendix

Detection and quantitation for mosaicism

Detection

DHPLC (denaturing high performance liquid chromatography)
MS (mass spectrometry)
RFLP (restricted fragment length polymorphisms)
SSCP/HD (single-strand conformation polymorphism/heteroduplex analysis)
PTT (protein truncation test)

Properties

Unable to quantify the fraction of mutant alleles
Detection limit as low as 0.5%

Quantitation

Pyrosequencing
Mismatch-high resolution melting curve
TaqMan assays based **digital PCR** or allele-specific qPCR **(TaqMAMA)**
MIP (molecular inversion probes), **RCA** (rolling circle amplification) based NGS approaches
Ultra deep sequencing: **MDS** (maximum-depth sequencing) and **duplex sequencing**

Properties

Special designed assay required
Limit: cost-effective approaches only have 5%-10% detection limit, those with detection limit lower than 0.1% were not affordable for large-scale screening

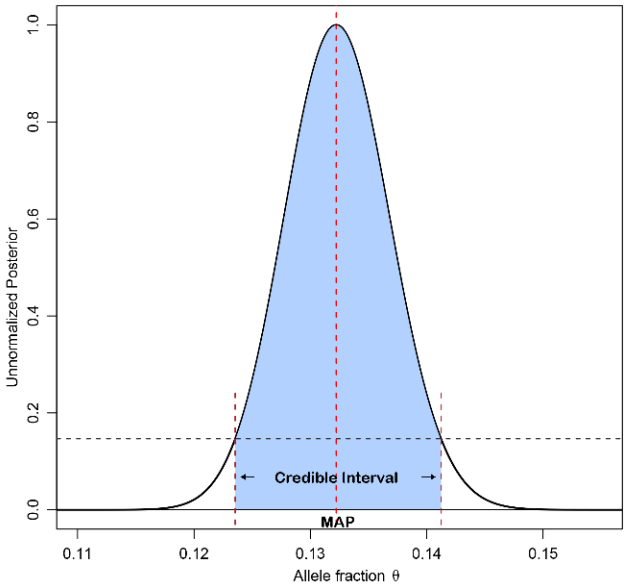
Large-scale quick quantification of low-freq mosaicism is required for distinguish mosaicism from "*de novo*" mutations

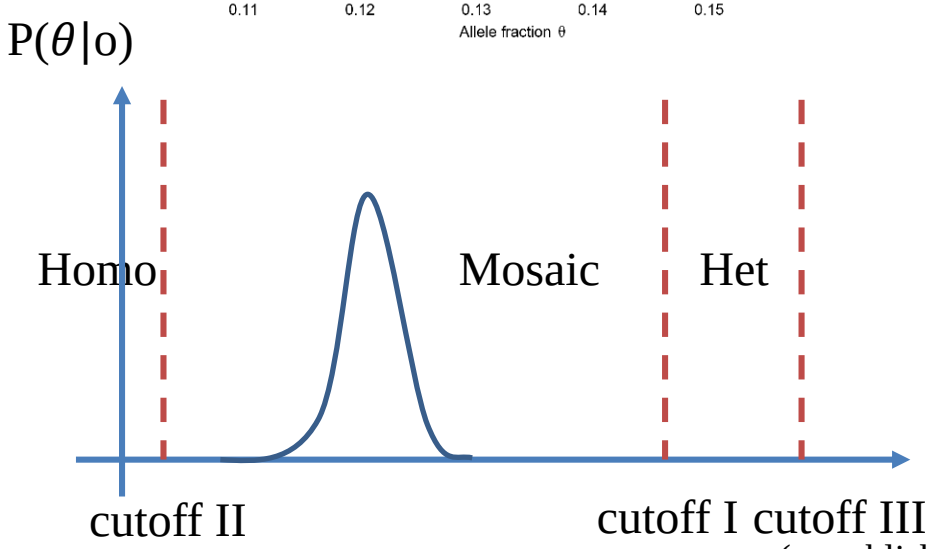
Appendix Bayesian model for calculating the fraction of mutant alleles

θ: theoretical fractions of the mutant alleles
o: number of reads support mutant alleles
n: the total number of reads mapped to the position
r: unobserved "real" number of allele count
$\mathrm{P}(\theta)$: prior
$\mathrm{P}(r|\theta;n)$: the likelihood of Bernoulli sampling
$\mathrm{P}(o|r;q)$: the summarized probabilities

$$\mathrm{P}(\theta|o) \propto \mathrm{P}(\theta)\mathrm{P}(o|\theta)$$
$$= \mathrm{P}(\theta)\sum_r \mathrm{P}(o,r|\theta)$$
$$= \mathrm{P}(\theta)\sum_r \mathrm{P}(r|\theta)\,\mathrm{P}(o|r)$$
$$= \mathrm{P}(\theta)\sum_r \mathrm{P}(r|\theta;n)\,\mathrm{P}(o|r;q)$$

Equation 0





(unpublished data)

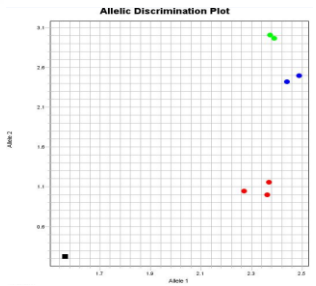
Appendix Micro-droplet digital PCR, **mDDPCR** for benchmarking PASM

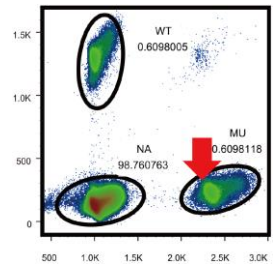
Target region

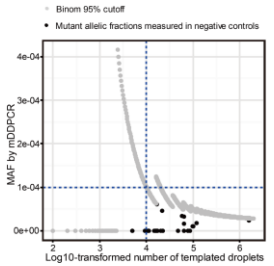
Design probe-based genotyping qPCR assays

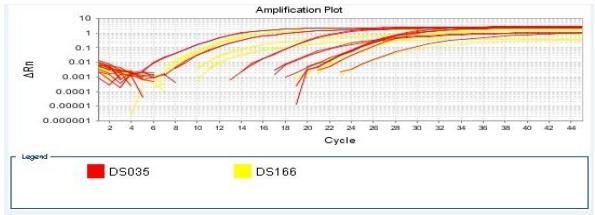
Data analysis and calculation of the 95% binominal confidence intervals

QC for the genotyping assay with qPCR end-point genotyping method





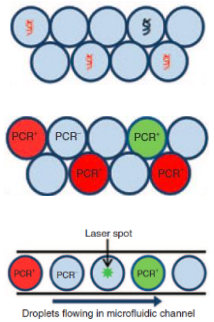




Fragmentation of DNA samples through sonication

Micro droplet generation, PCR amplification and micro-droplet detection

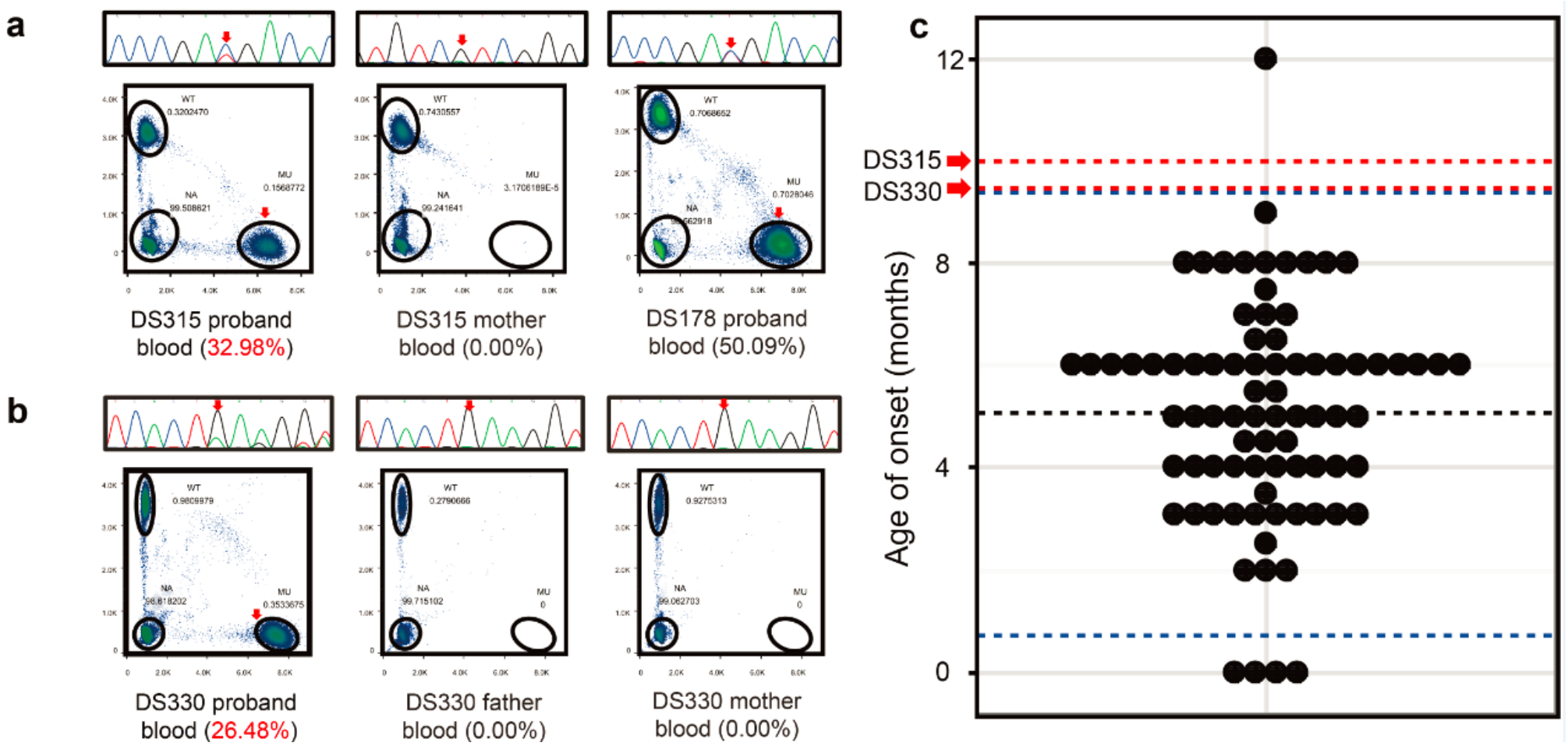
(Under review)



http://covarisinc.com/products/afa-ultrasonication/m-series/

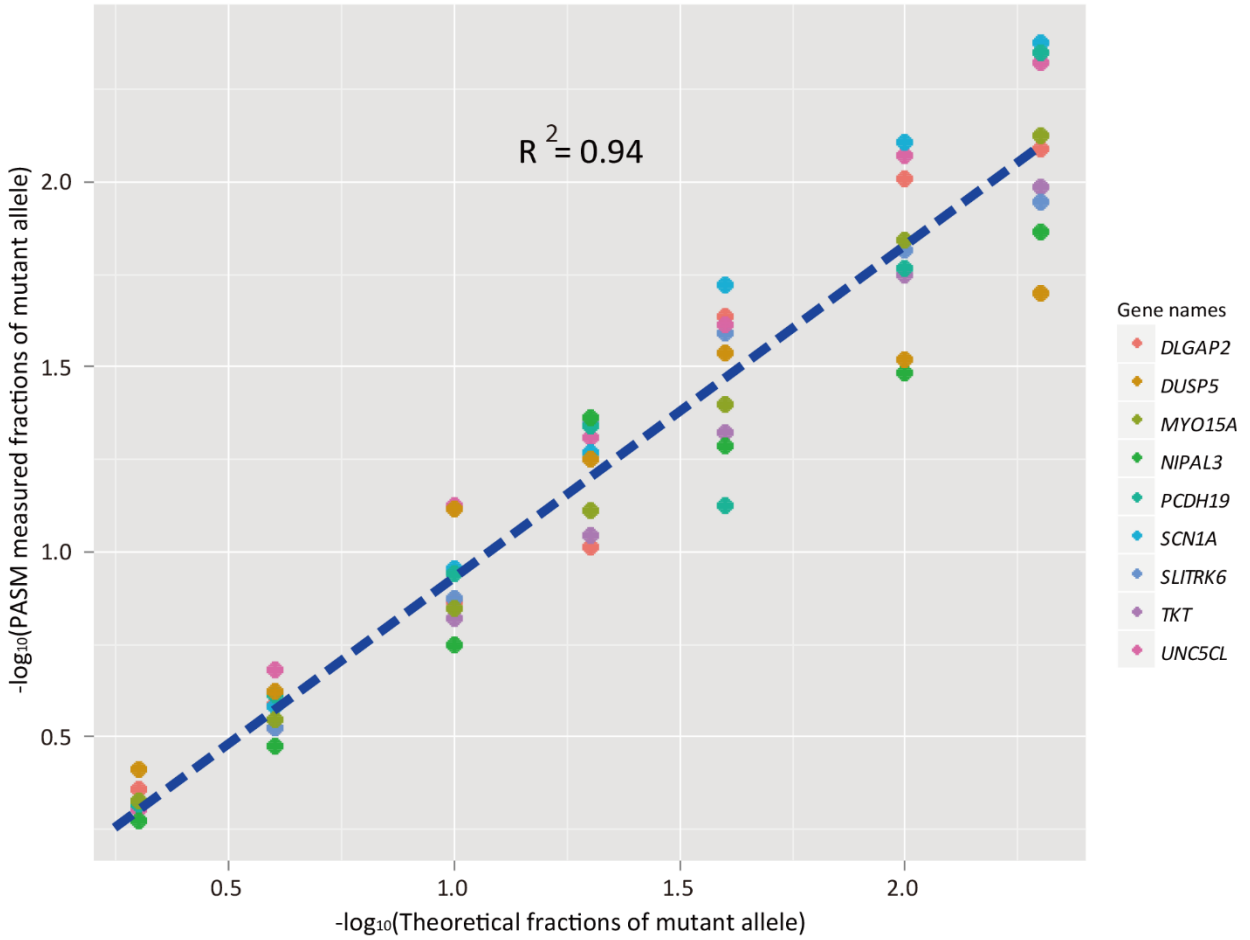
(Modified from Chen, *et al*., *Mol Ther Nucleic Acid*s, 2013)

Appendix Genotype and phenotype of patient mosaicism in *SCN1A* mutated DS affected families



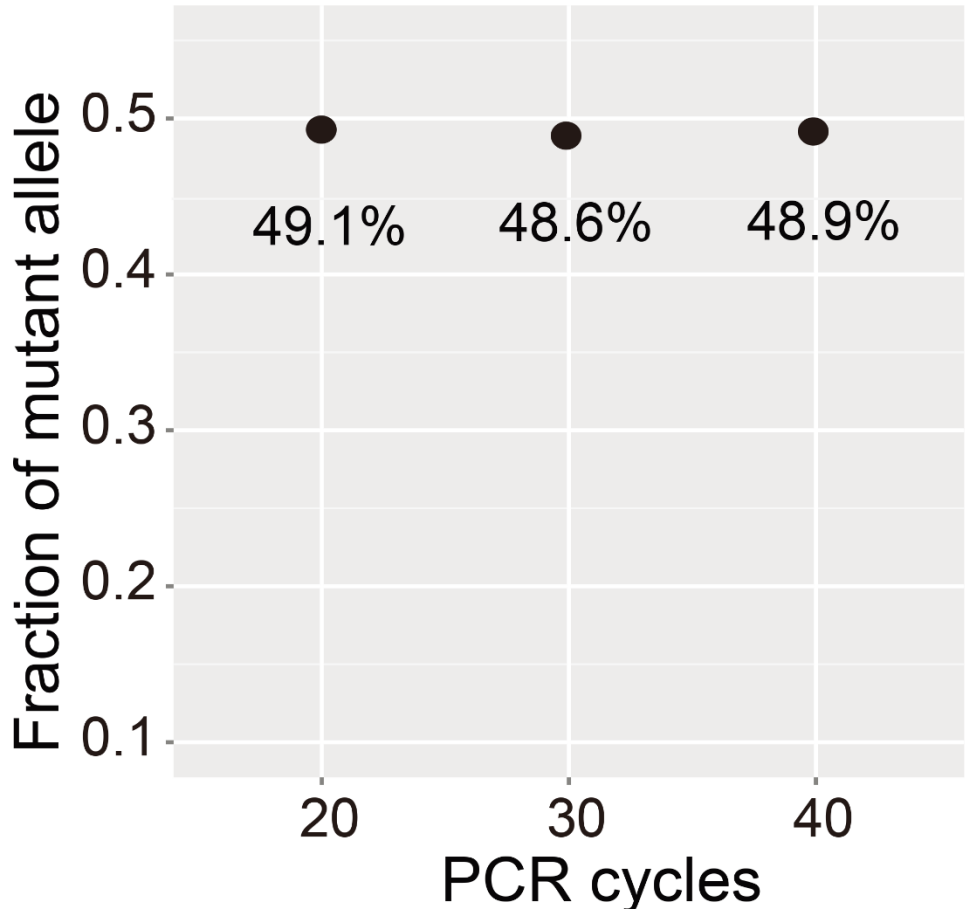
(revision submitted)

Appendix Benchmark PASM with standards and replicates



Unpublished data)

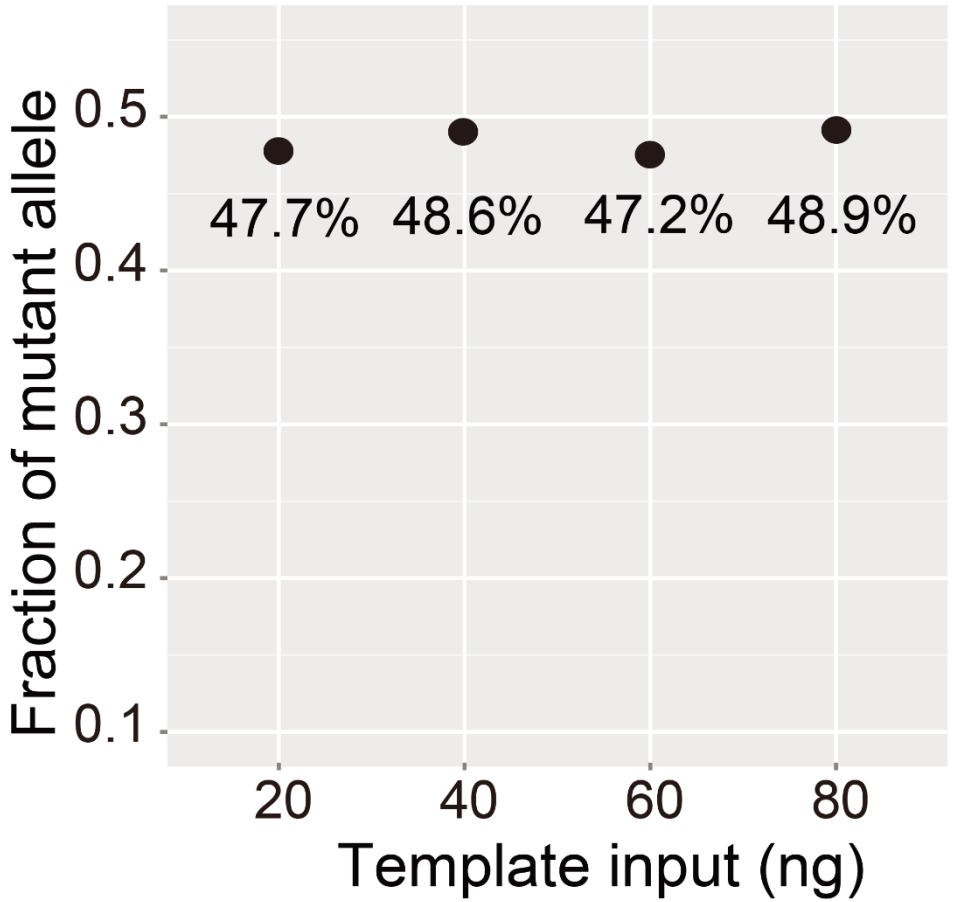
Appendix Stability of PASM with different PCR cycles



(Xu, *et al.*, *Hum Mutat*, 2015)

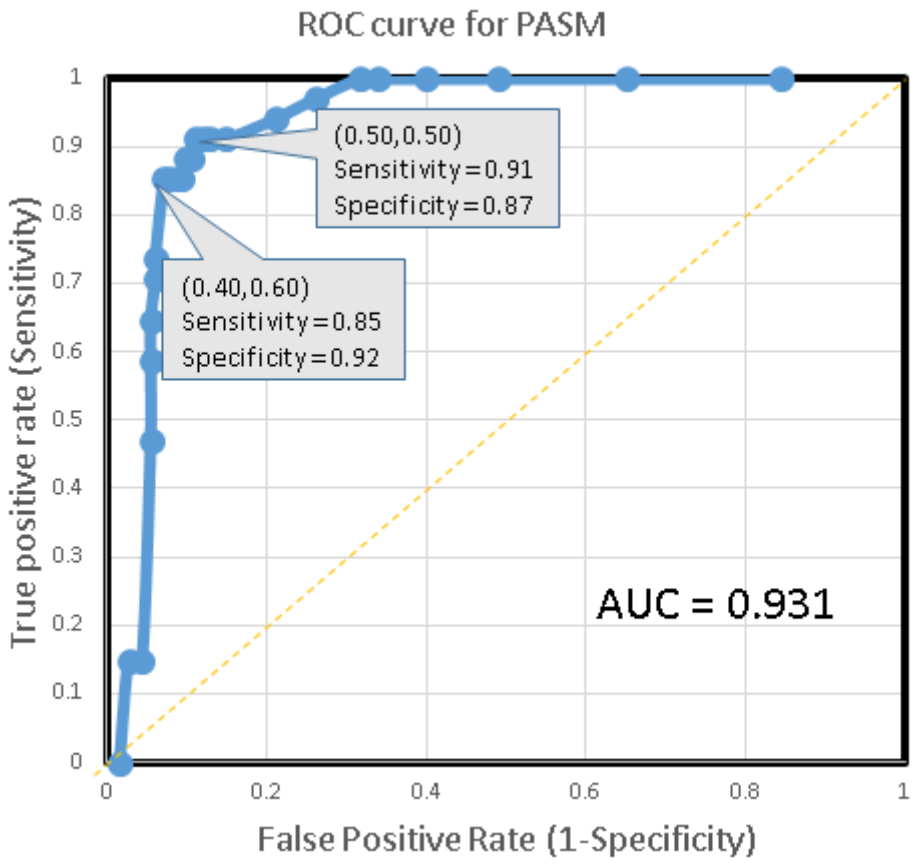
Appendix

Stability of PASM with different amounts of input template DNA



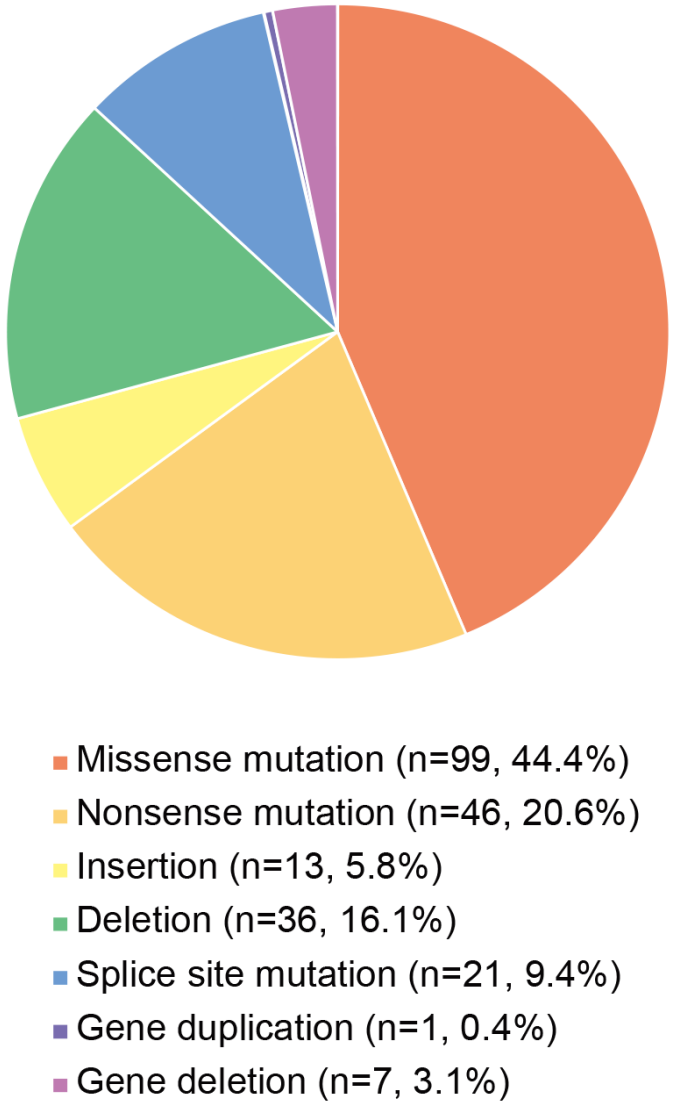
(Xu, *et al.*, *Hum Mutat*, 2015)

Appendix ROC of PASM, use mDDPCR as gold standard



(unpublished data)

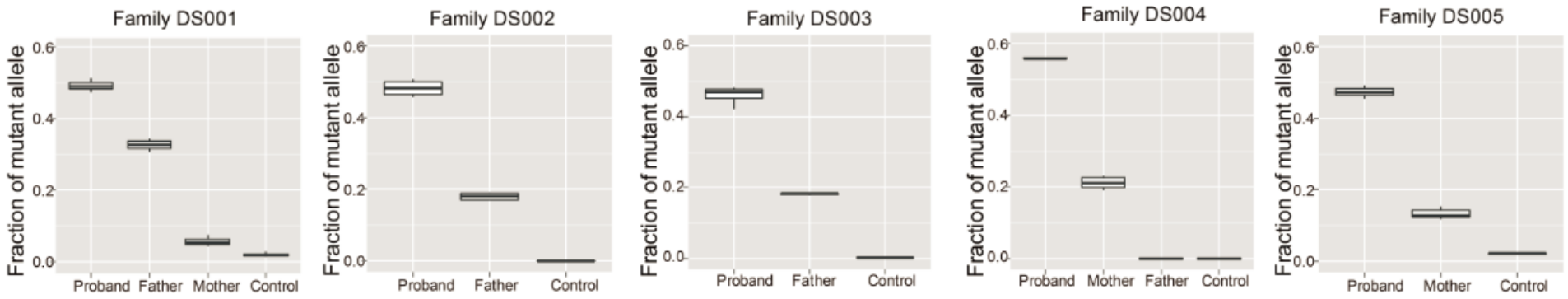
Appendix Types of functional mutations in *SCN1A* in the DS probands



(Xu, *et al.*, *Hum Mutat*, 2015)

Appendix

Quantification of Sanger sequencing detectable mosaicism by PASM

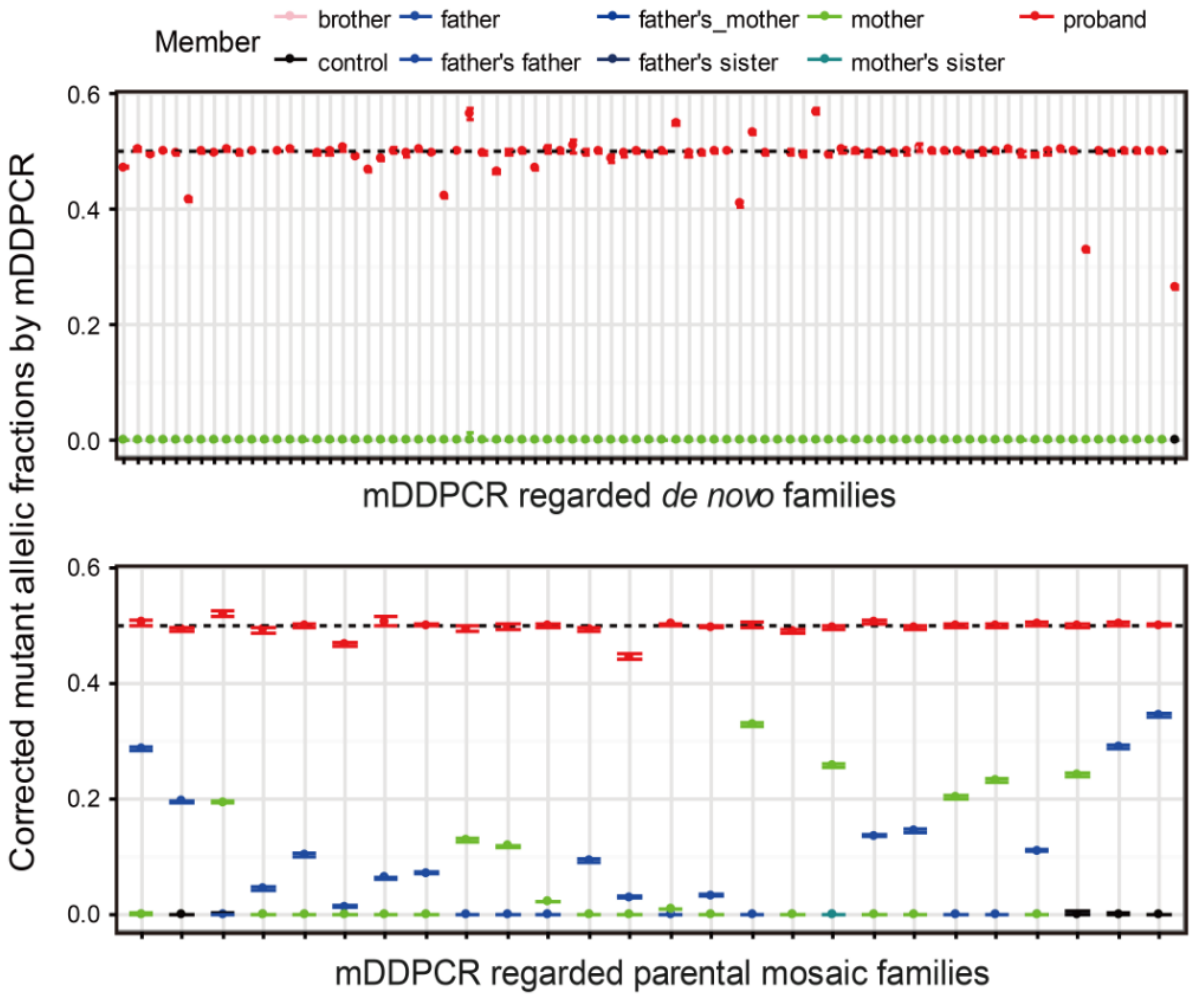


		Fractions of mutated allele measured by PASM in peripheral blood			
Number	Variants	Proband	Mother	Father	Negative control
DS001	c.1118 del T	49.19±1.92%	5.64±1.62%	32.61±1.88%	0.79±0.16%
DS002	c.4351C>A	48.30±2.40%		18.05±1.05%	0.03±0.00%
DS003	c.2593 C>T	46.10±2.65%		18.21±0.32%	0.17±0.02%
DS004	c.5003C>G	56.07±0.37%	21.15±1.86%	0.00±0.01%	0.00±0.01%
DS005	c.4302G>A	47.36±2.55%	13.26±1.42%		0.28±0.26%

(Xu, *et al.*, *Hum Mutat*, 2015)

Appendix

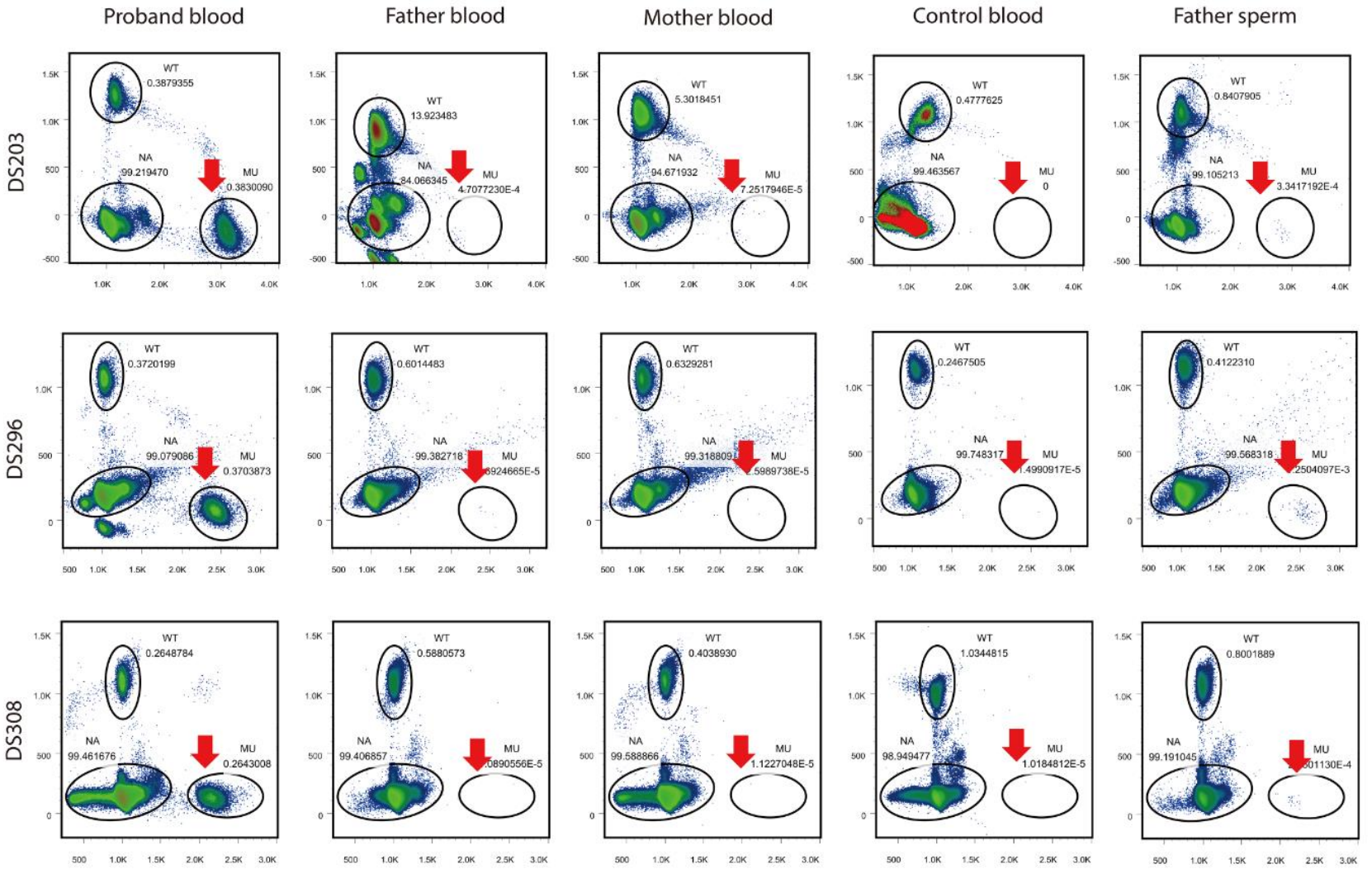
26 parental mosaicisms were detected from the blood sample of 112 *SCN1A* mutated DS families by mDDPCR



(under review)

Appendix

Three paternal mosaicisms were only detectable in the father's **sperm**



(under review)

Appendix

Conditional probability correction for the MAF elevation in paternal sperm

$$P_{sp|T} = \frac{P_{sp,T}}{P_T} = \frac{P_{T|sp} \cdot P_{sp}}{P_{T|sp} \cdot P_{sp} + P_{T|bl} \cdot P_{bl}} = \frac{MAF_{large} \cdot 0.5}{MAF_{large} \cdot 0.5 + MAF_{small} \cdot 0.5} = \frac{MAF_{large}}{MAF_{large} + MAF_{small}}$$

P_T: probability of transmitting a deleterious mutation to a child
P_{sp}: probability of observing an MAF in sperm higher than blood
P_{bl}: probability of observing an MAF in sperm lower than blood
MAF_{large}: the larger of the MAFs measured in paternal blood and sperm
MAF_{small}: smaller of the MAFs measured in paternal blood and sperm

Equation 1

$$P_{corrected} = \prod_{i=1}^{n} P_{i,sp|T} = \prod_{i=1}^{n} \frac{MAF_{i,large}}{MAF_{i,large} + MAF_{i,small}}$$

Equation 2

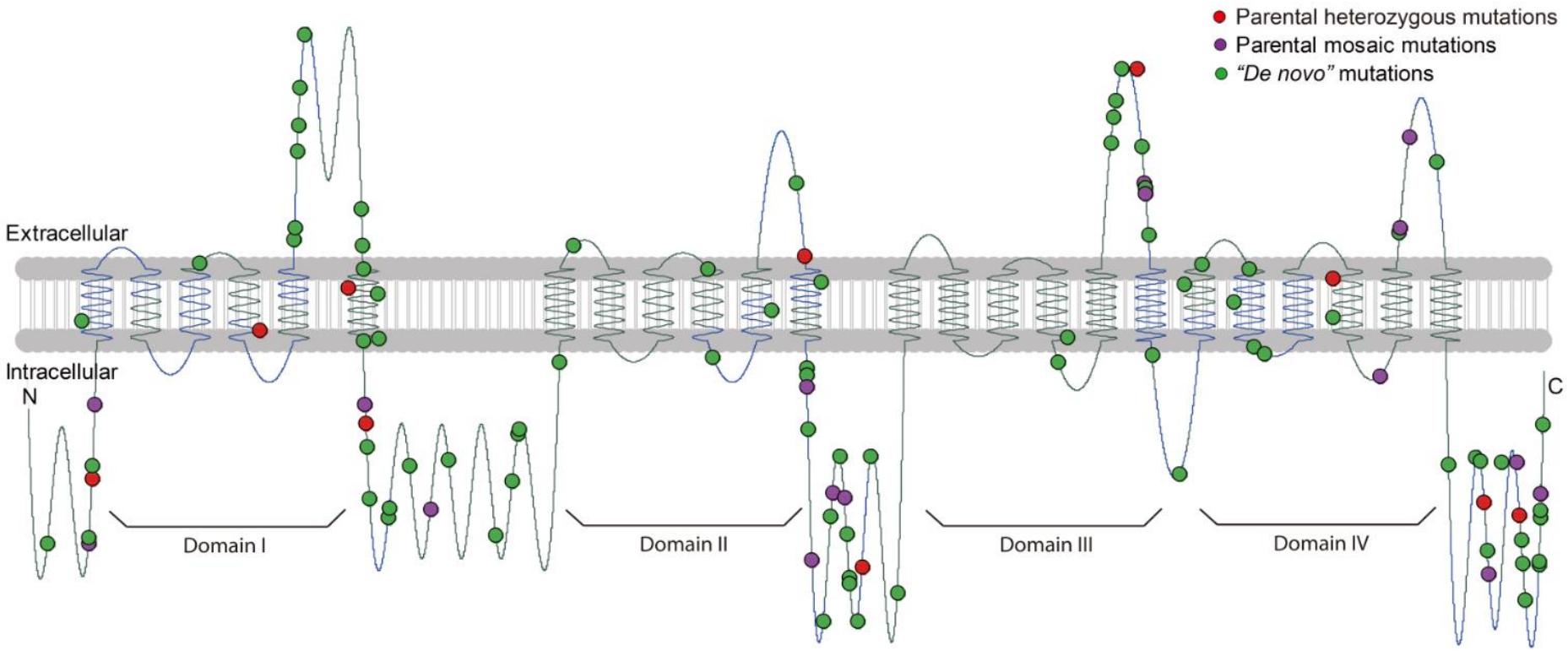
After conditional probability correction,

P' = 0.033

(revision submitted)

Appendix

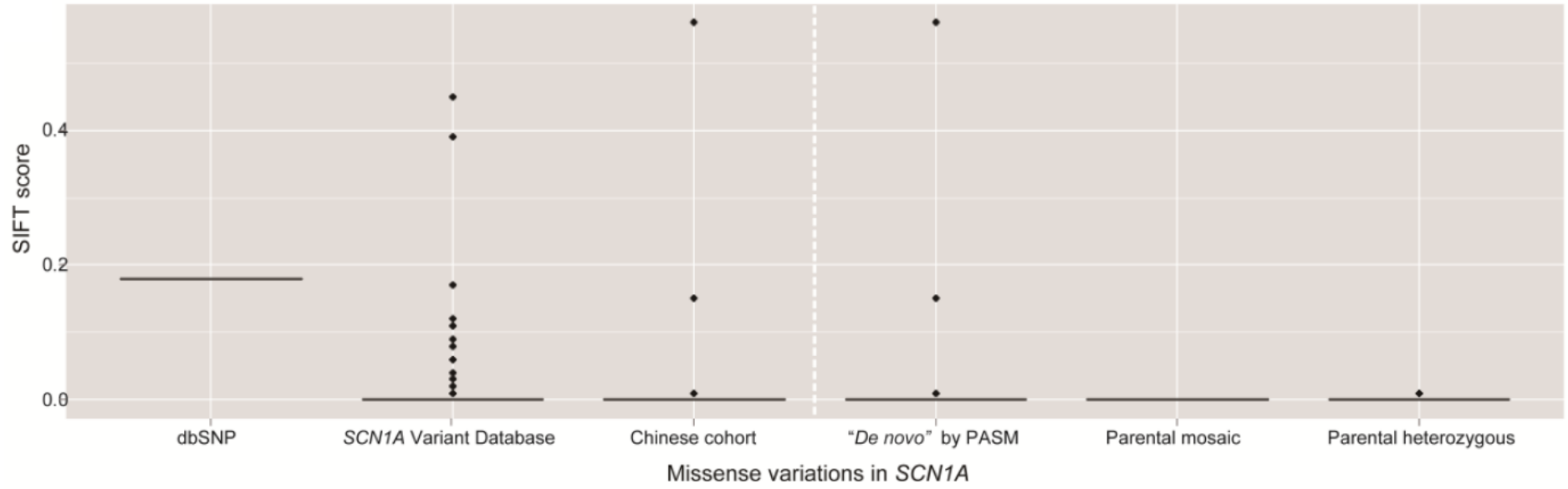
Locations of the mutations on the domain structures of SCN1A protein



(Xu, *et al.*, *Hum Mutat*, 2015)

Appendix

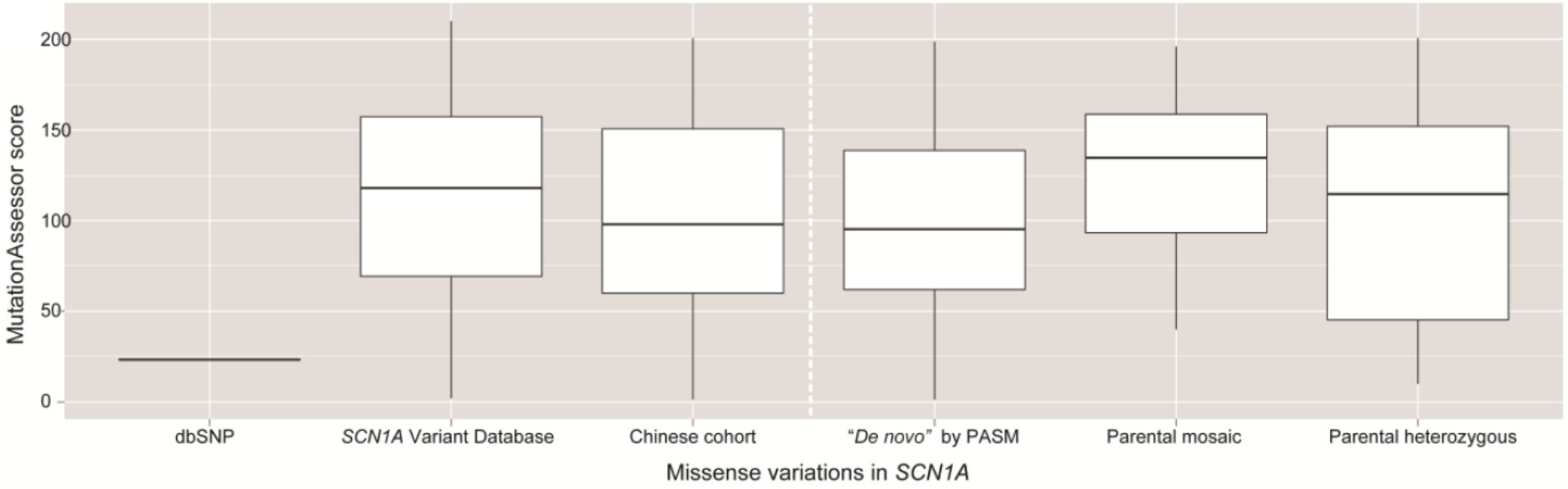
Functional predictions of *SCN1A* mutations by SIFT



(Xu, *et al.*, *Hum Mutat*, 2015)

Appendix

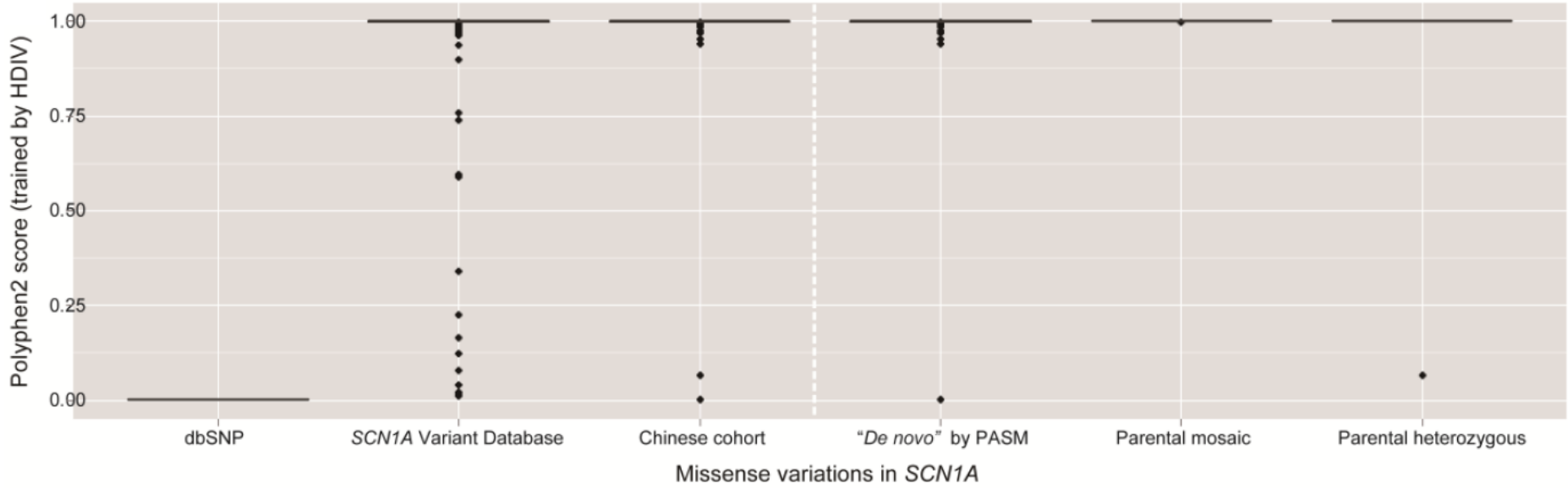
Functional predictions of *SCN1A* mutations by MutationAssessor



(Xu, *et al.*, *Hum Mutat*, 2015)

Appendix

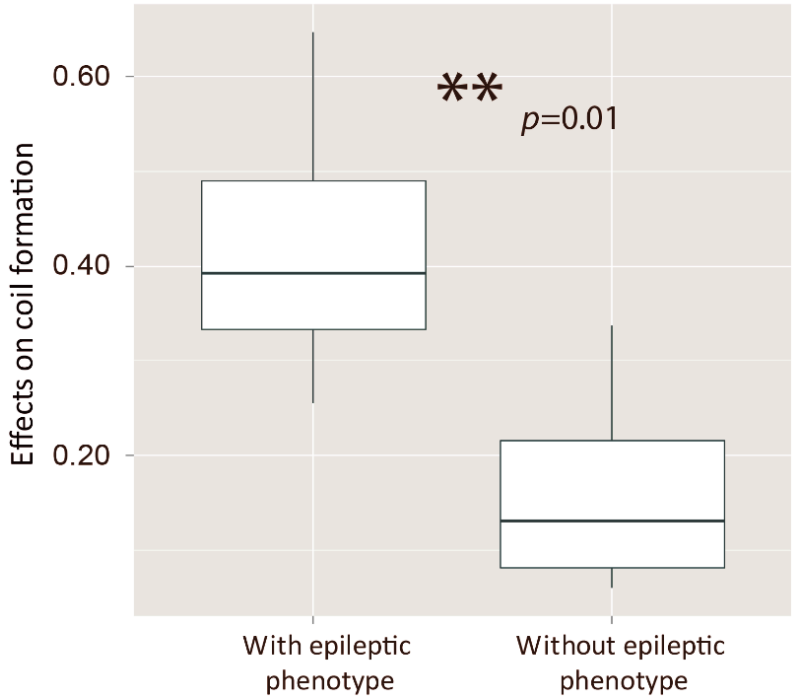
Functional predictions of *SCN1A* mutations by Polyphen2 (HDIV)



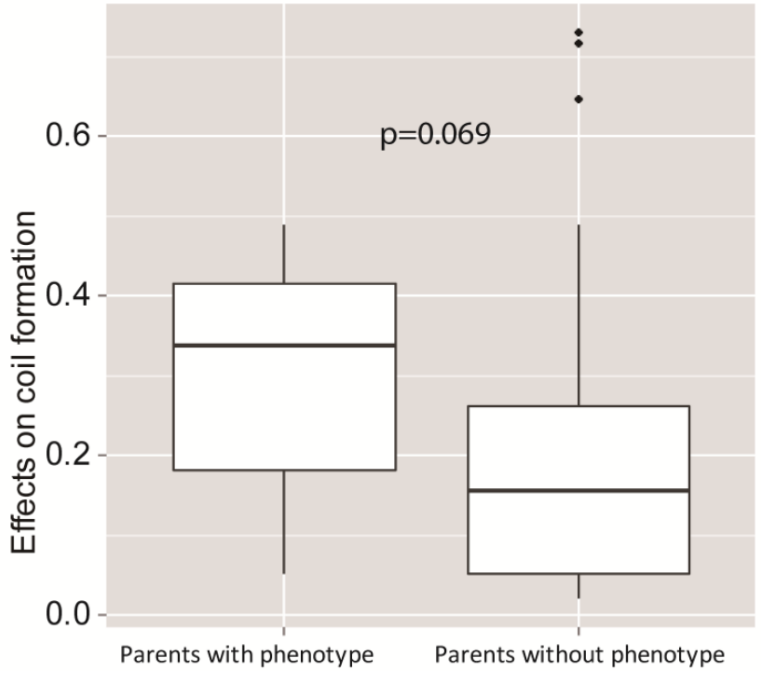
(Xu, *et al.*, *Hum Mutat*, 2015)

Appendix

Functional differences of *SCN1A* mutations by Protscale



Mosaic mutations by PASM



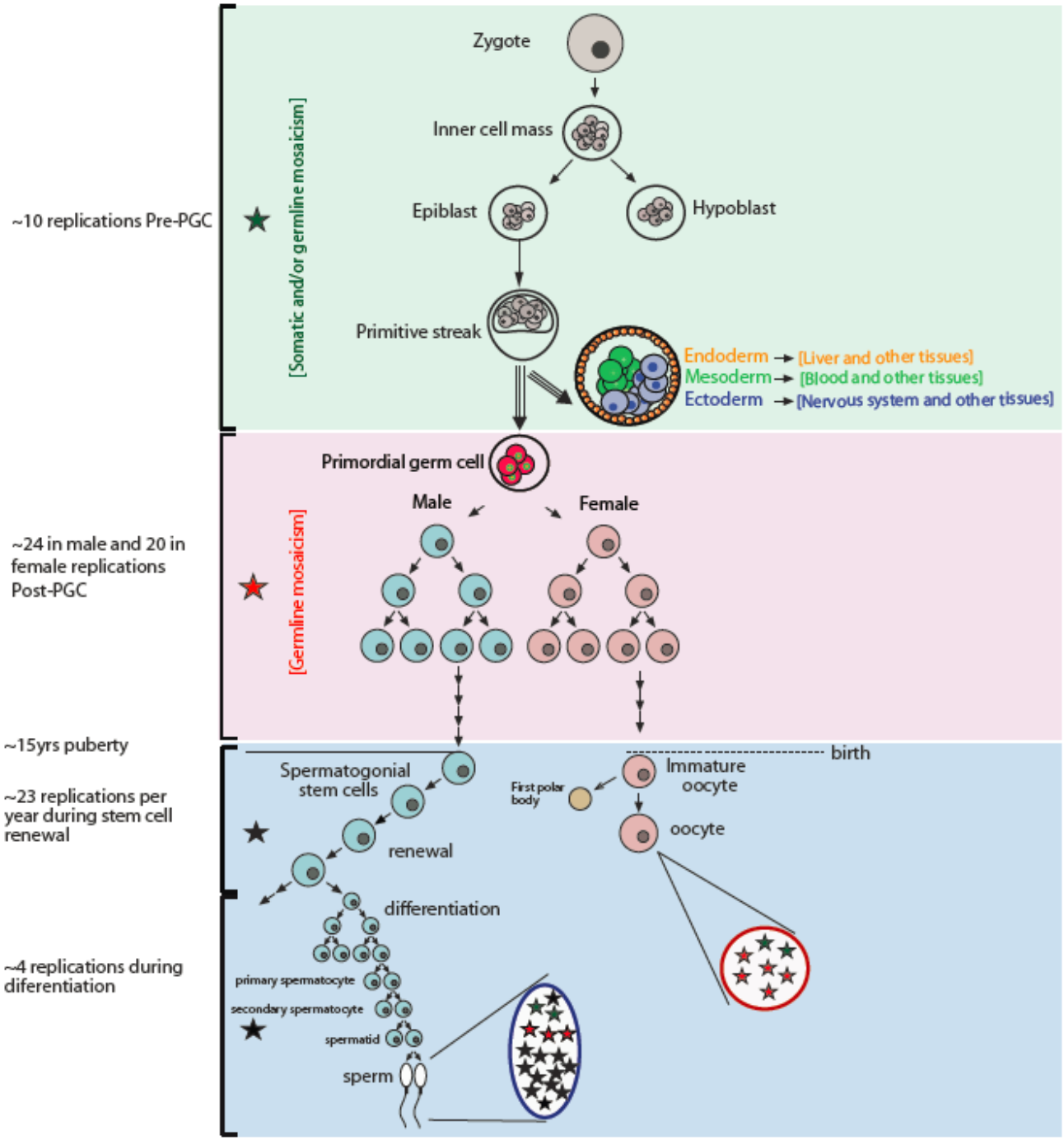
"*De novo*" mutations by PASM

(Xu, *et al.*, *Hum Mutat*, 2015)

Appendix

Possible explanation for the higher MAF in sperm than blood

Appendix

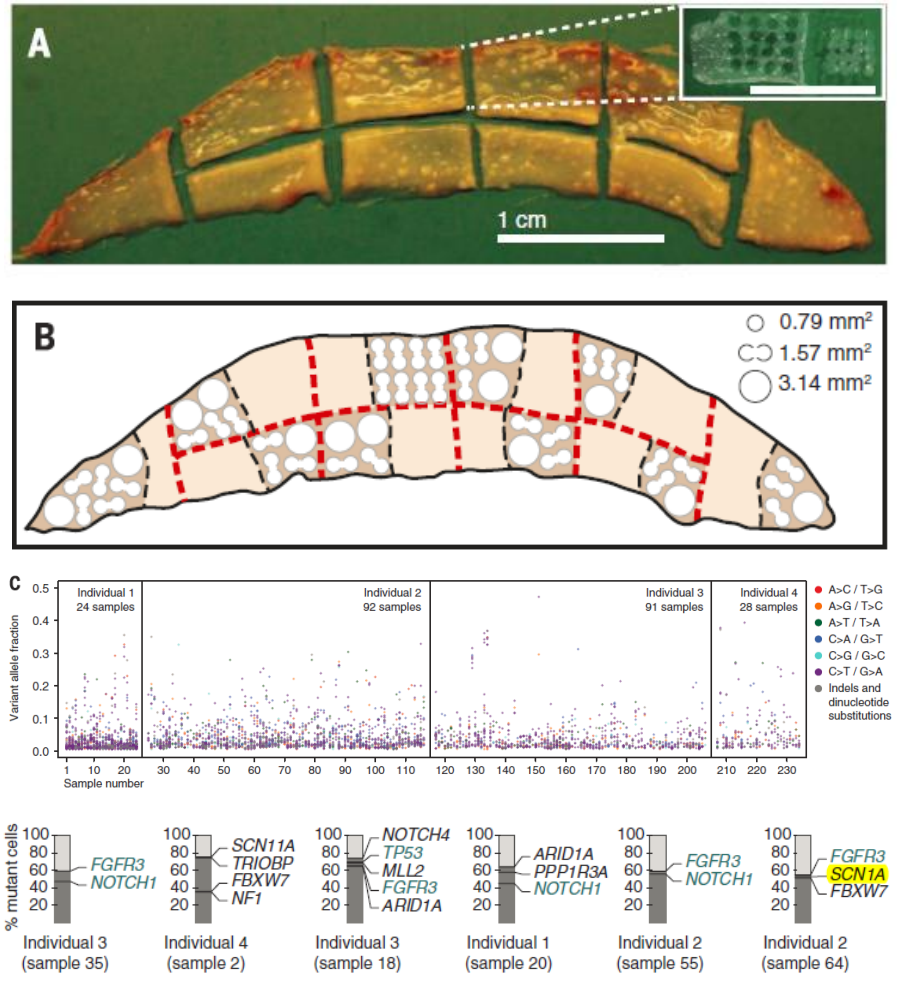


(Rahbari *et al.*, *Nat Genet*, 2015)

Appendix

SCN1A is on the top list of positive selected somatic mutations in normal human skin

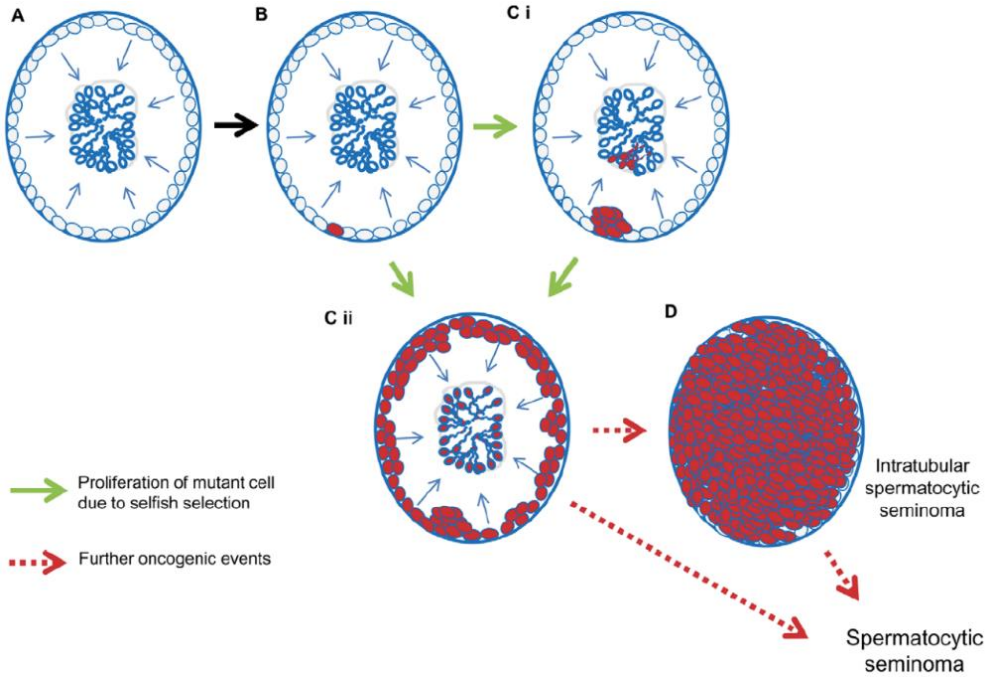
Rank	Gene name	Number of variants
1	NOTCH1	451
2	NOTCH2	194
3	FAT1	147
4	MUC17	147
5	FAT4	143
6	APOB	133
7	MLL2	102
8	SPHKAP	102
9	PREX2	92
10	SCN1A	92
11	TP53	91
12	BAI3	90
13	ERBB4	89
14	GRM3	85
15	PPP1R3A	81
16	SALL1	81
17	NOTCH3	74
18	SCN11A	72
19	ADAMTS18	62
20	GRIN2A	58



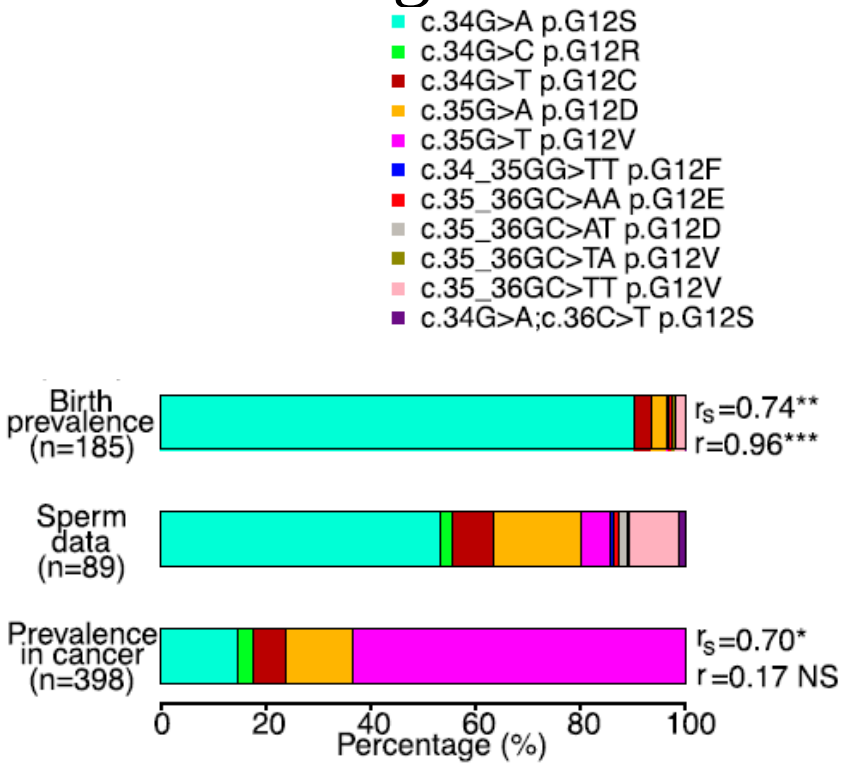
(Summarized from the supp data of Martincorena *et al.*, *Science*, 2015)

Appendix

Cell-level selection might explain the paternal age effect (PAE) and elevated mosaic MAFs in parental germline in disorders caused by cancer-related genes



(Lim, *et al.*, *PLoS One*, 2012)



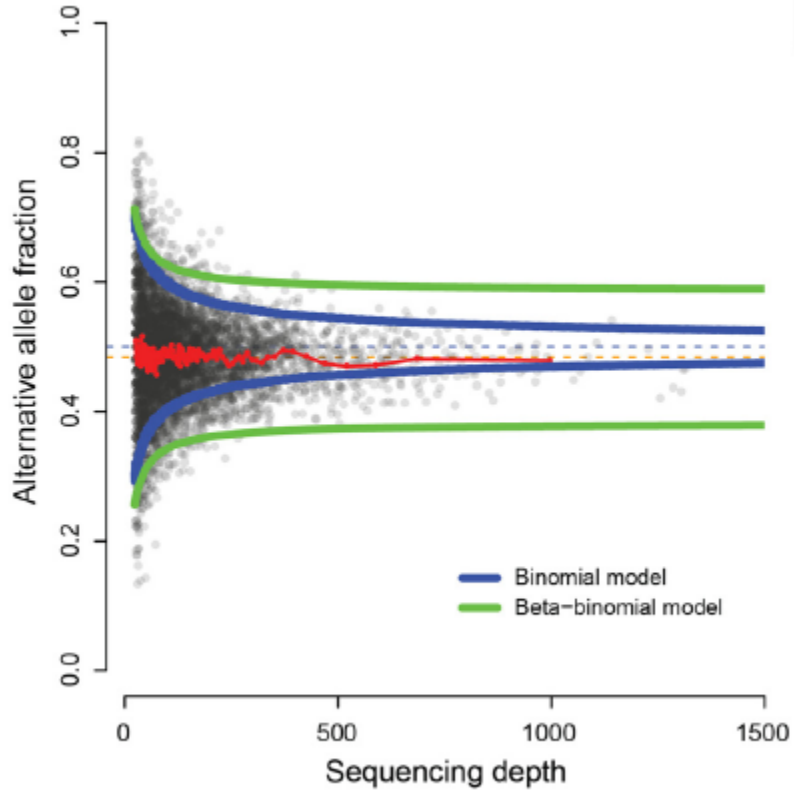
(Giannoulatou, *et al.*, *Proc Natl Acad Sci USA*, 2013)

Wilson *et al.* estimated that the male germ line has experienced 160 genome replications in a 20-year-old male and 610 genome replications in a 40-year-old male.

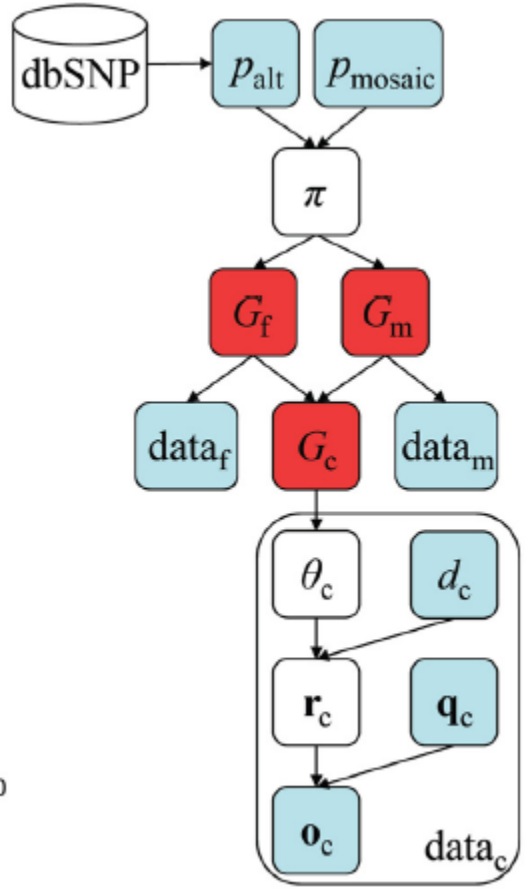
For disorders not caused by oncogenes, MAFs in sperm and blood remains largely unknown

Advantages of MosaicHunter

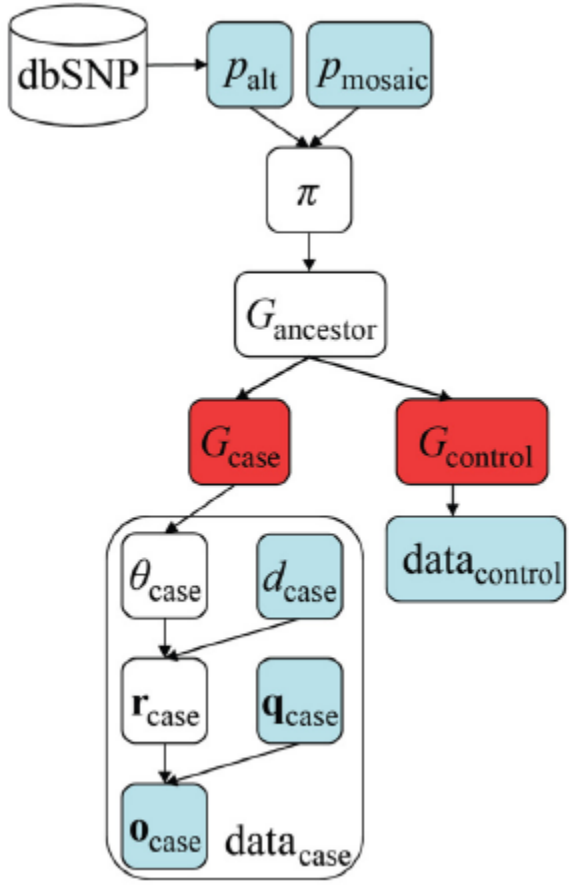
Appendix MosaicHunter incorporates various probabilistic models



Beta-binomial model for exome/captured data



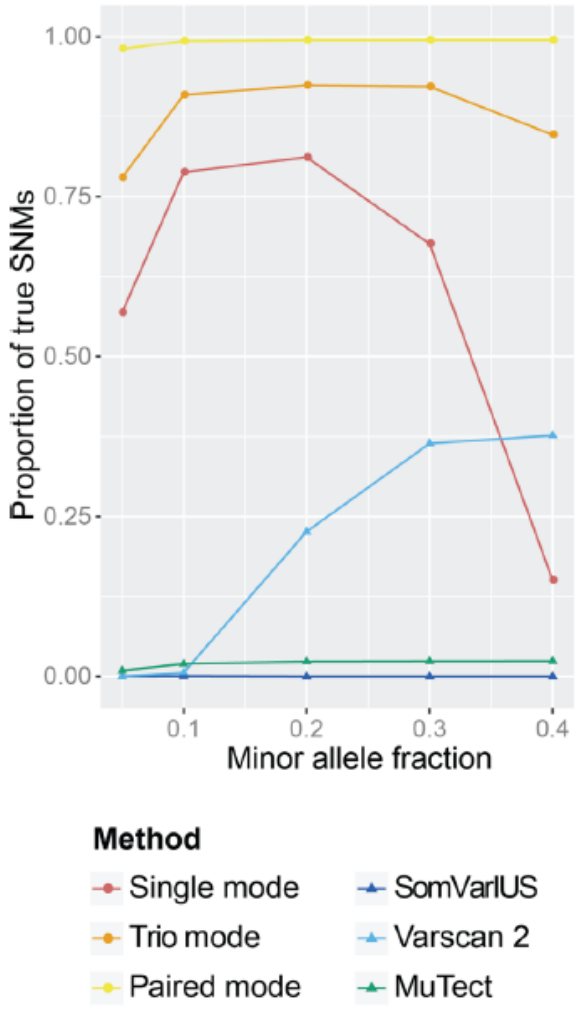
Trio Mode

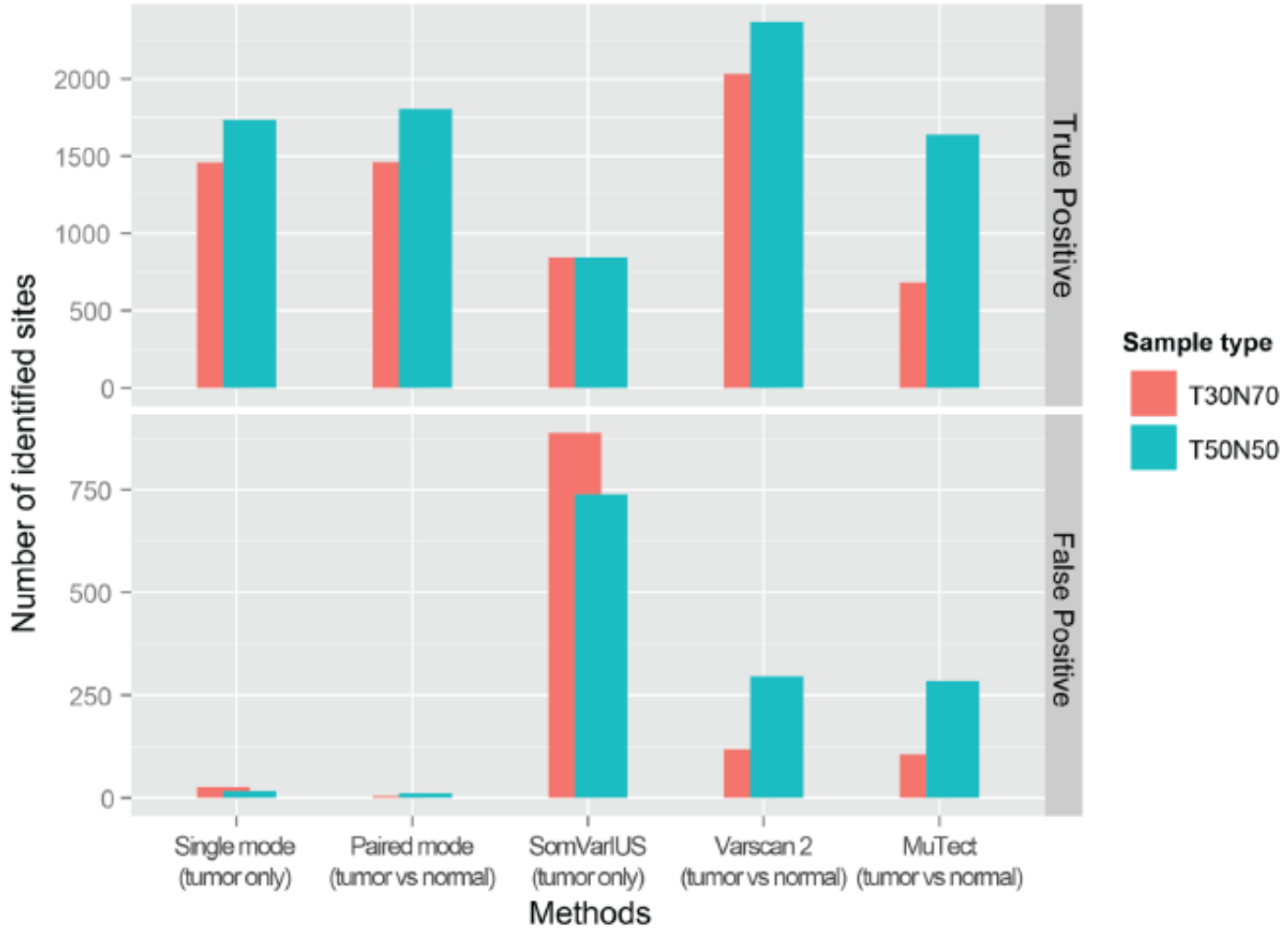


Paired Mode

(Huang, *et al.*, *Nucleic Acids Res.*, 2017)

MosaicHunter was benchmarked with very low false positive rate





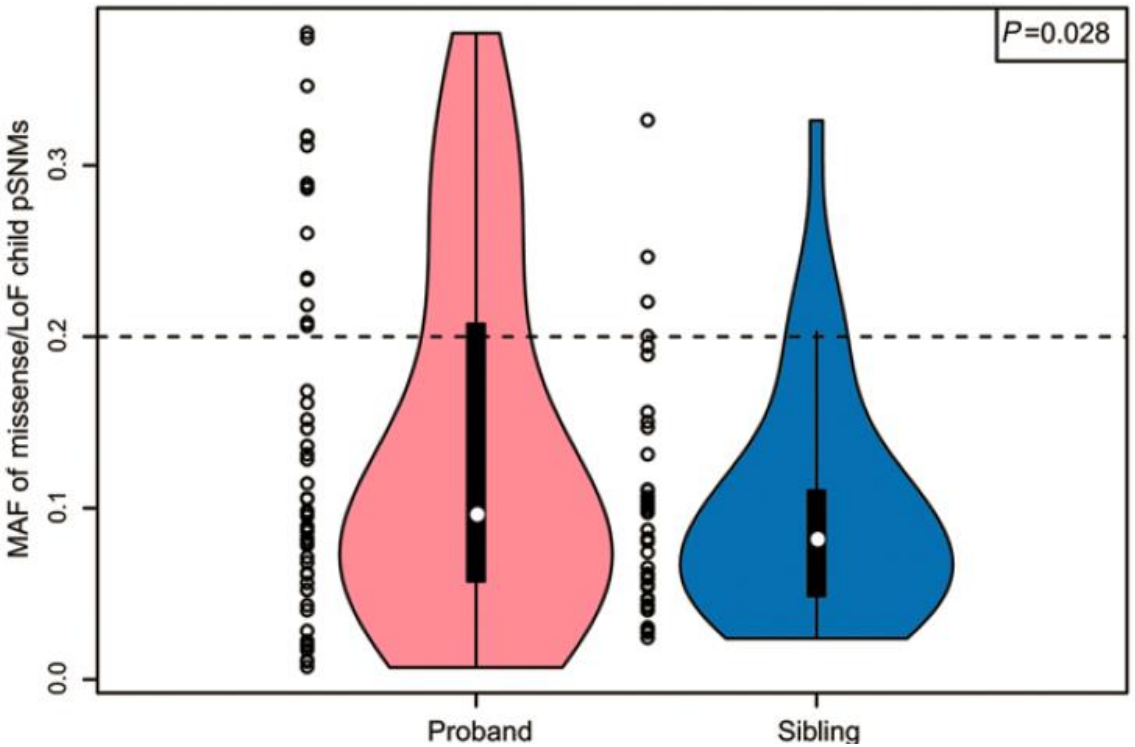
(Huang, *et al.*, *Nucleic Acids Res.*, 2017)

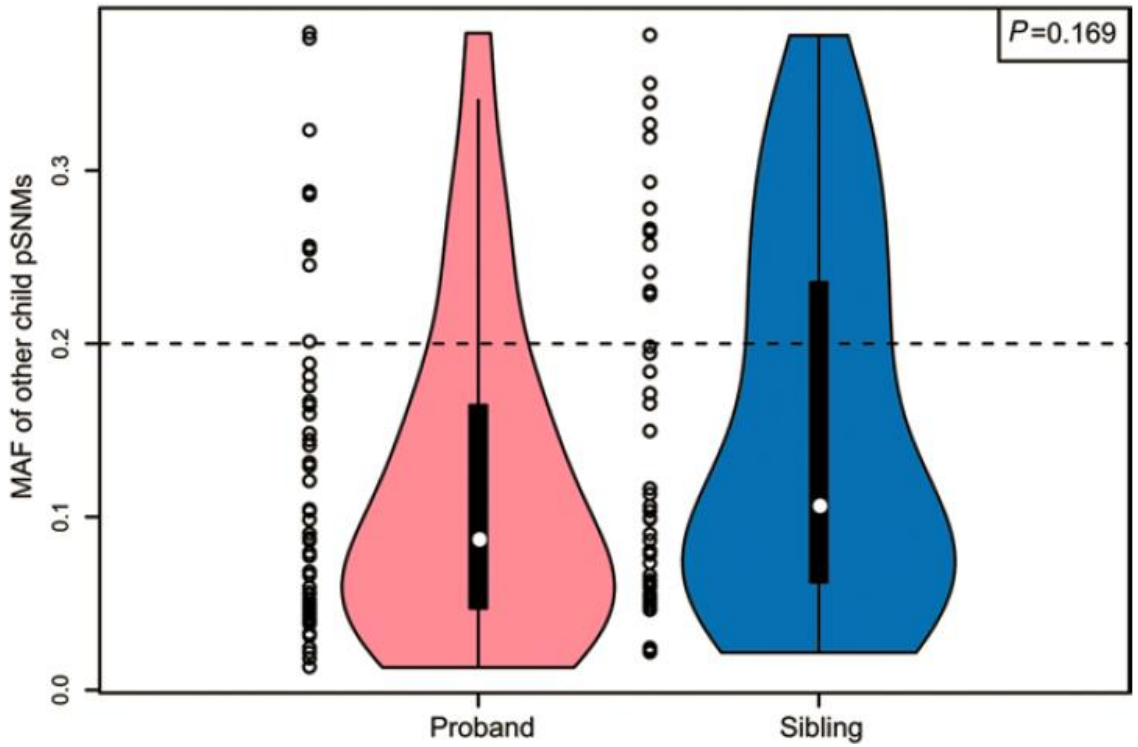
Appendix

pSNMs: burden and autistic phenotype

Appendix

MAF of missense/loss-of-function pSNMs were validated to be higher from ASD probands than from unaffected siblings

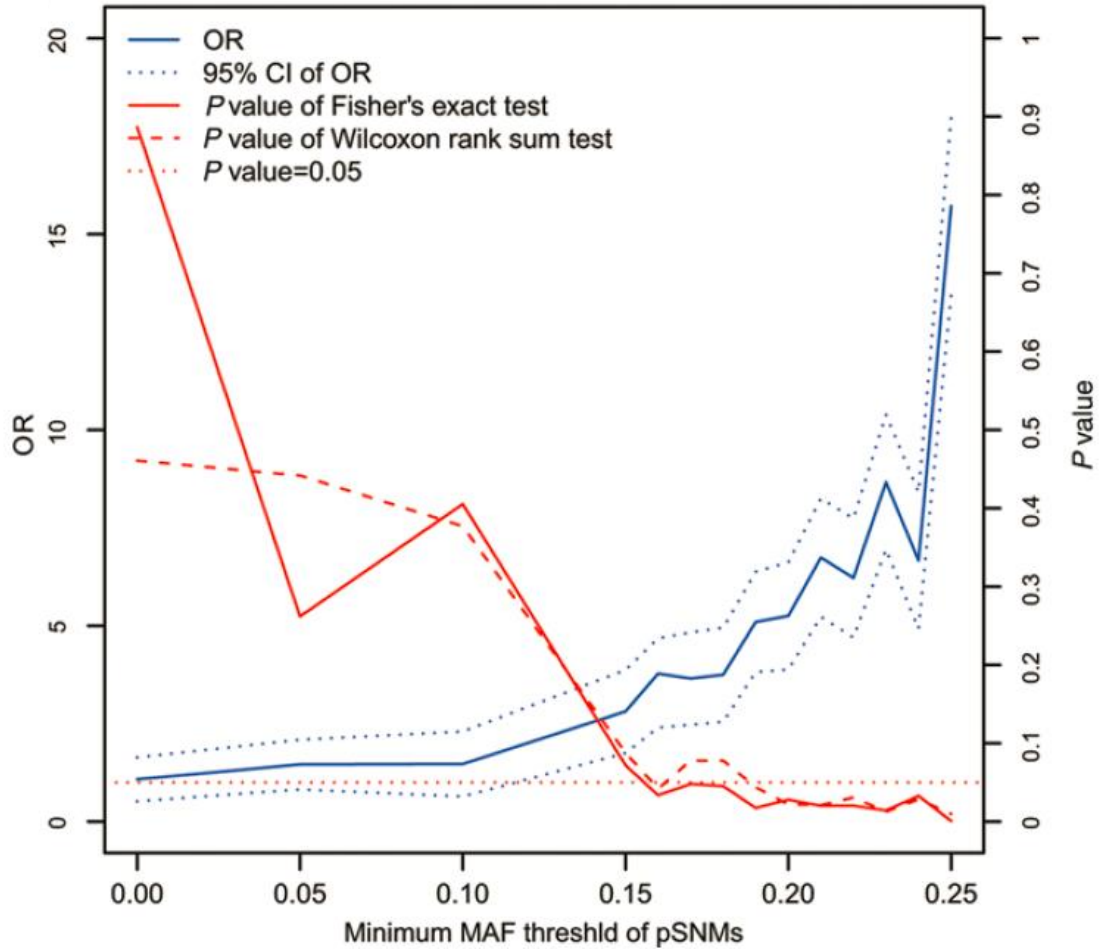


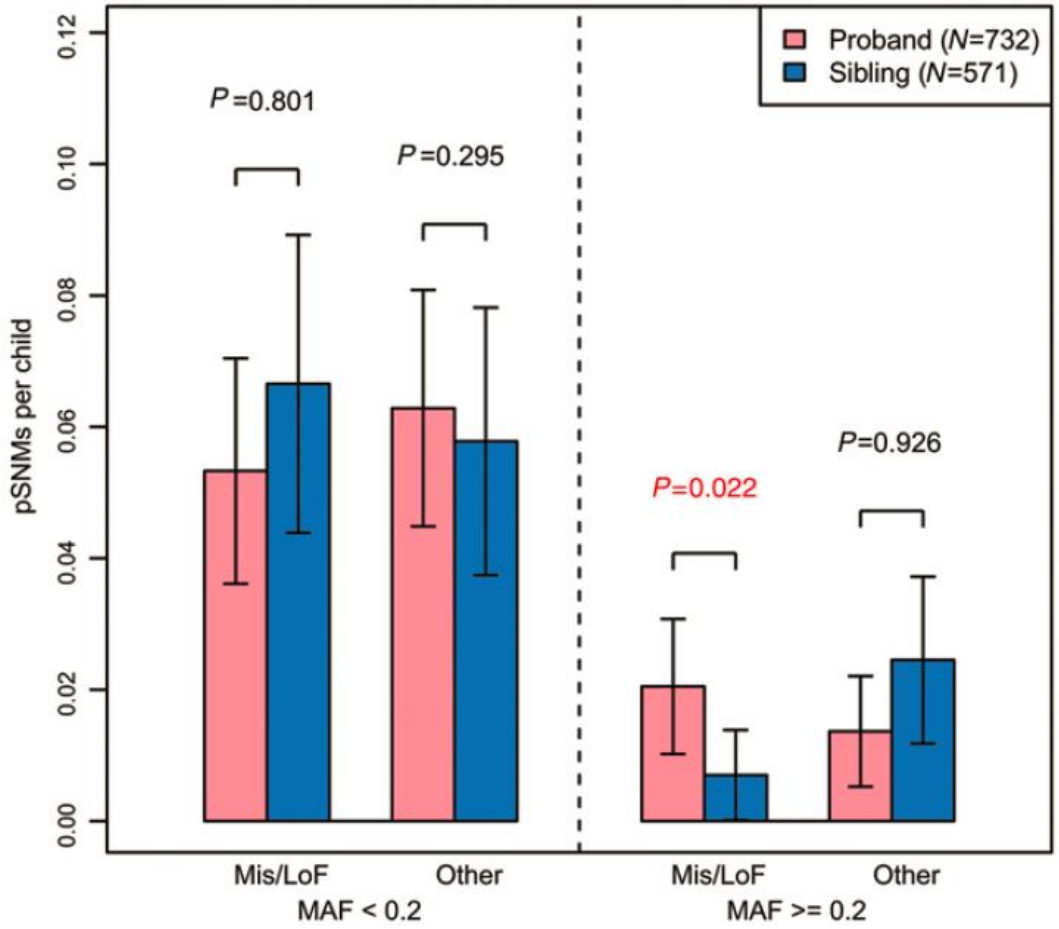


(Dou, *et al.*, *Hum Mutat*, 2017)

Appendix

Contribution of missense/Loss-of-Function child pSNMs with high mutant allele fractions to ASD diagnoses

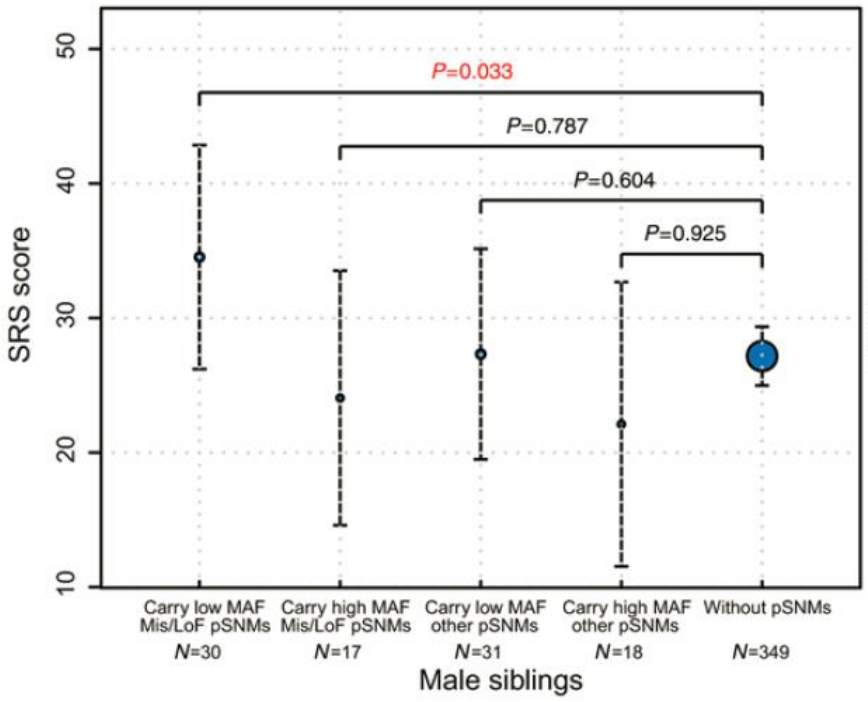


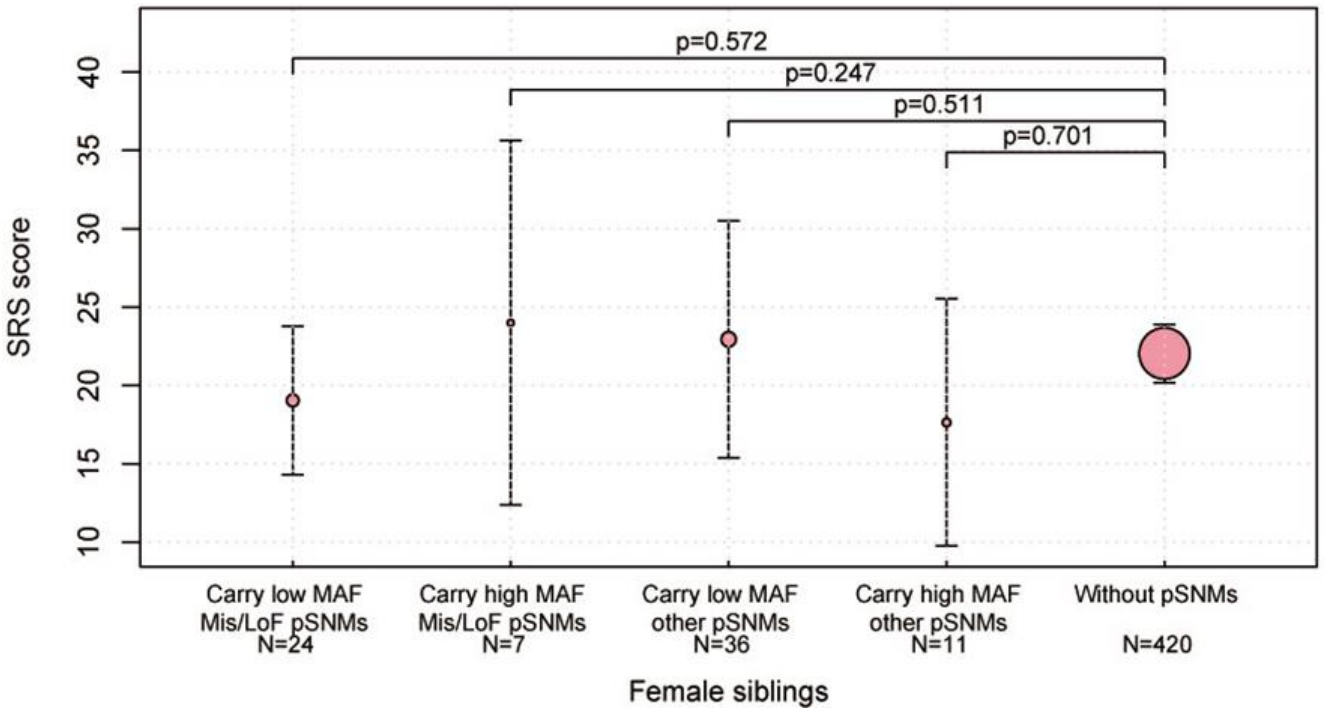


(Dou, *et al.*, *Hum Mutat*, 2017)

Appendix pSNMs with low MAFs contribute to autistic traits

SRS: Social Responsiveness Scale, a scoring system by the teacher to measure the autistic phenotype



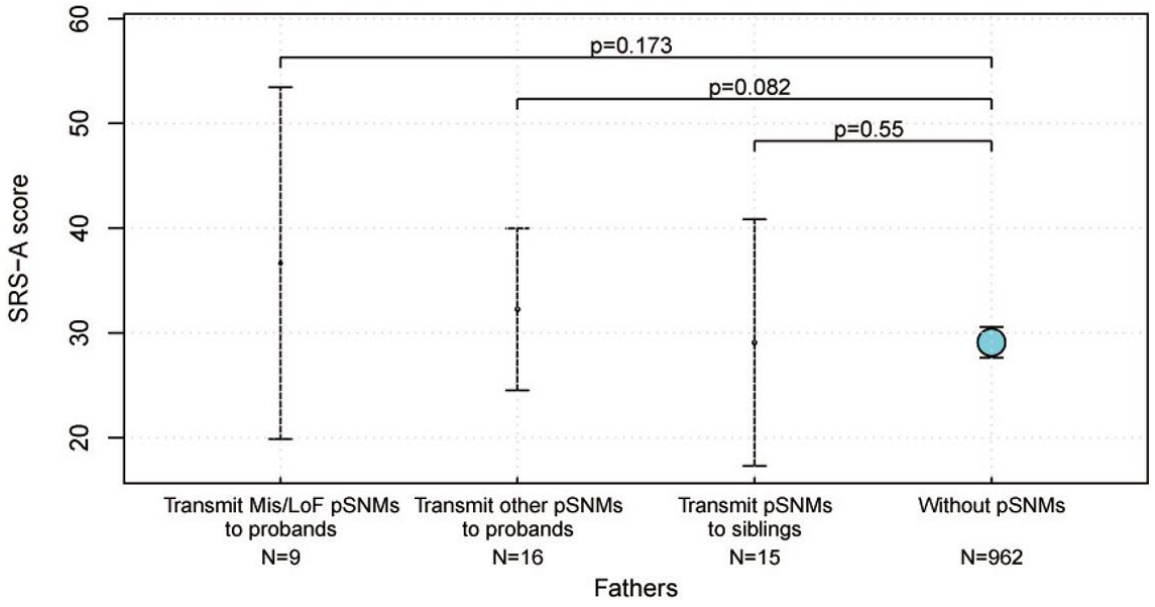


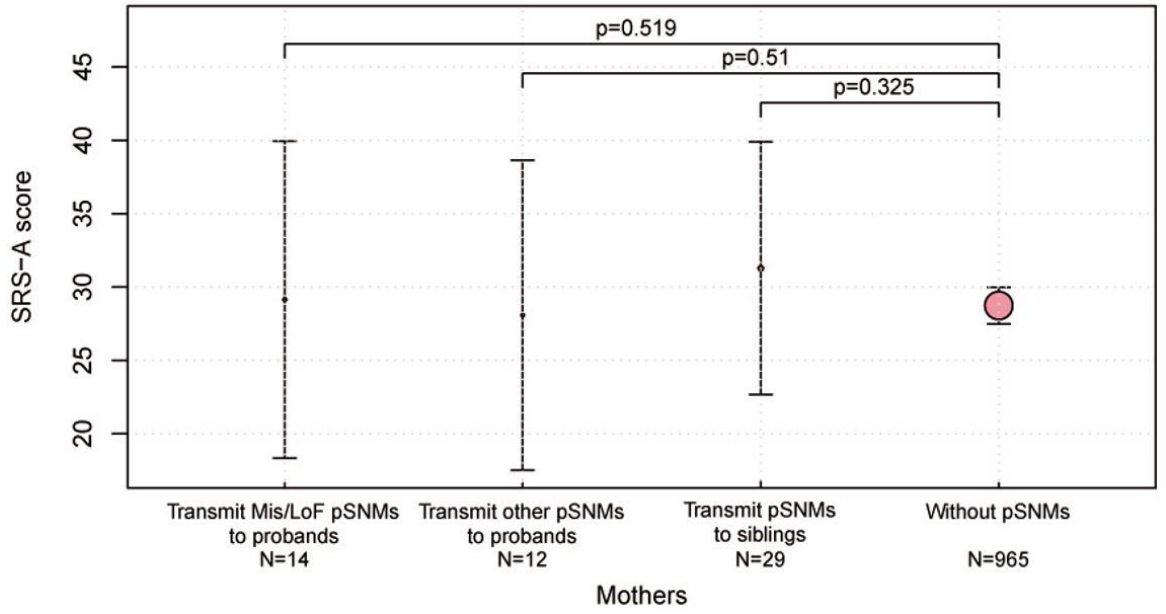
(Dou, *et al.*, *Hum Mutat*, 2017)

Appendix

Mosaic fathers transmitting missense/LoF pSNMs to probands seem to have higher SRS-A scores, the phenomenon was not observed in mothers

SRS-A: Social Responsiveness Scale for Adult, a scoring system to measure the autistic phenotype of ASD parents





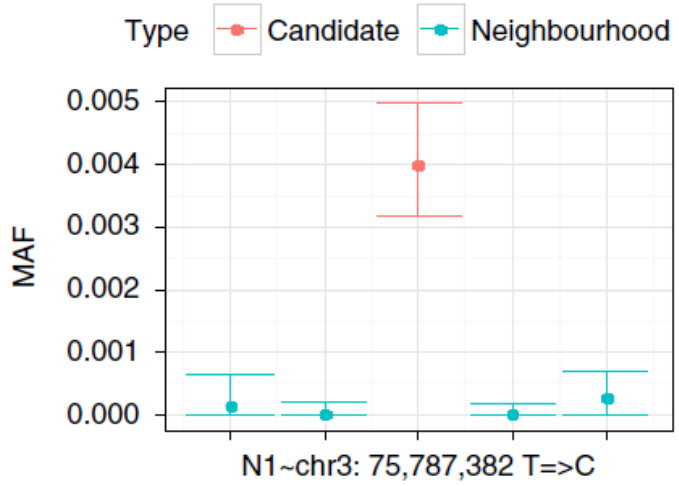
(Dou, *et al.*, *Hum Mutat*, 2017)

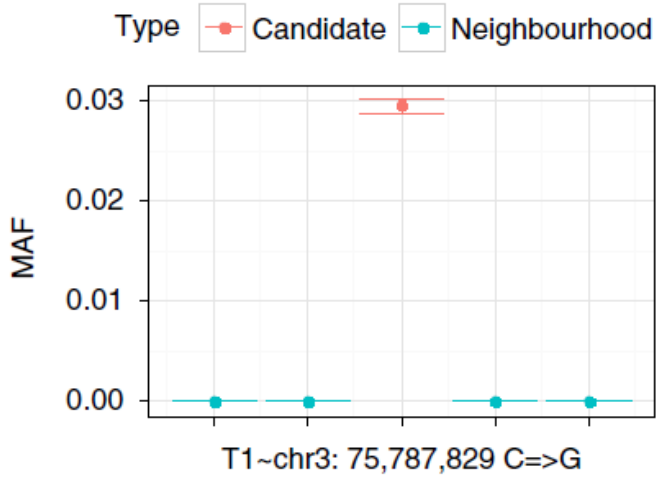
Appendix

Application of PASM and mDDPCR on ultra-low-fraction mosaic mutations in Cancer

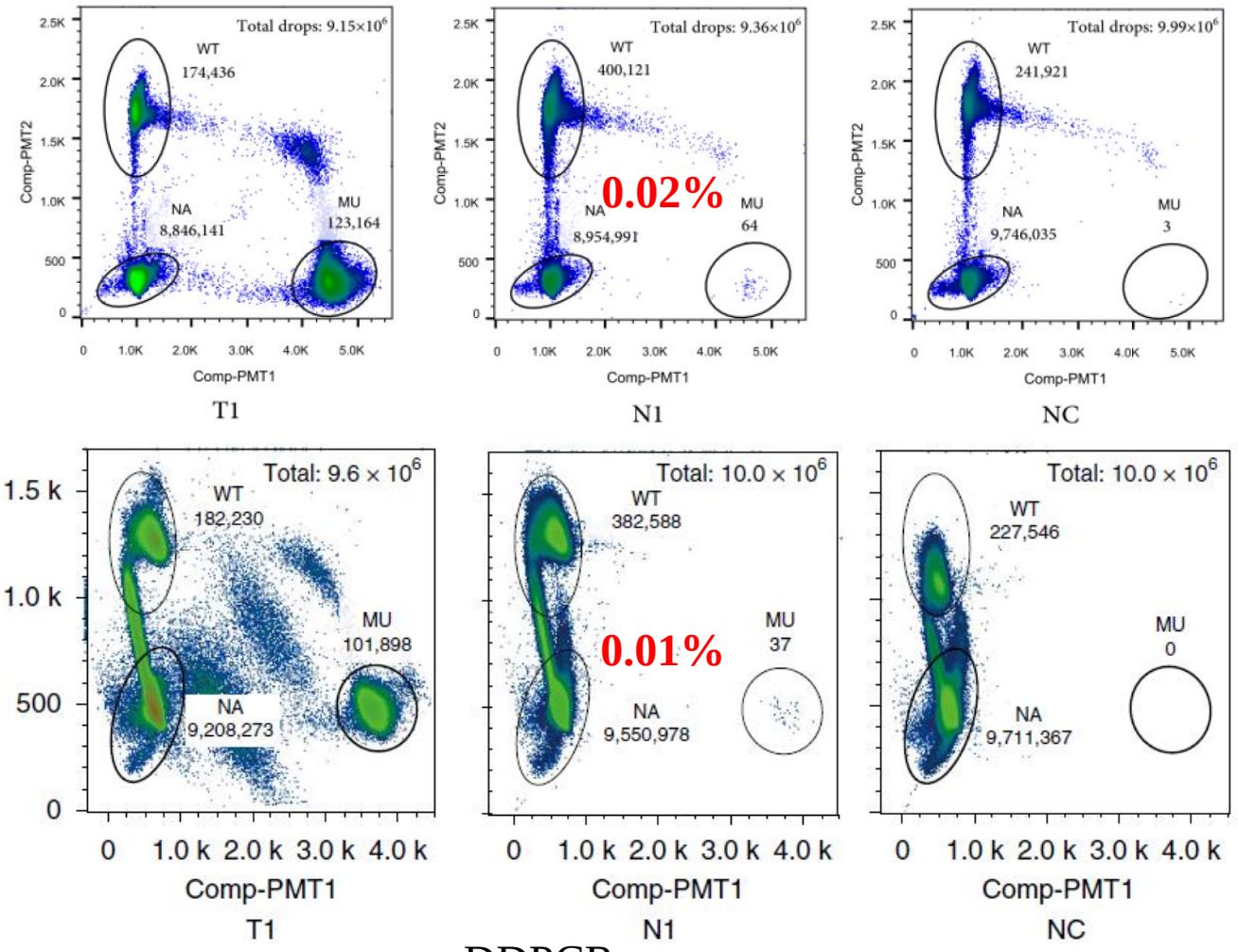
Appendix

PASM and mDDPCR were also powerful for ultra-low fraction mosaicism in cancer samples and validation of other NGS techniques





PASM+ Ampliseq

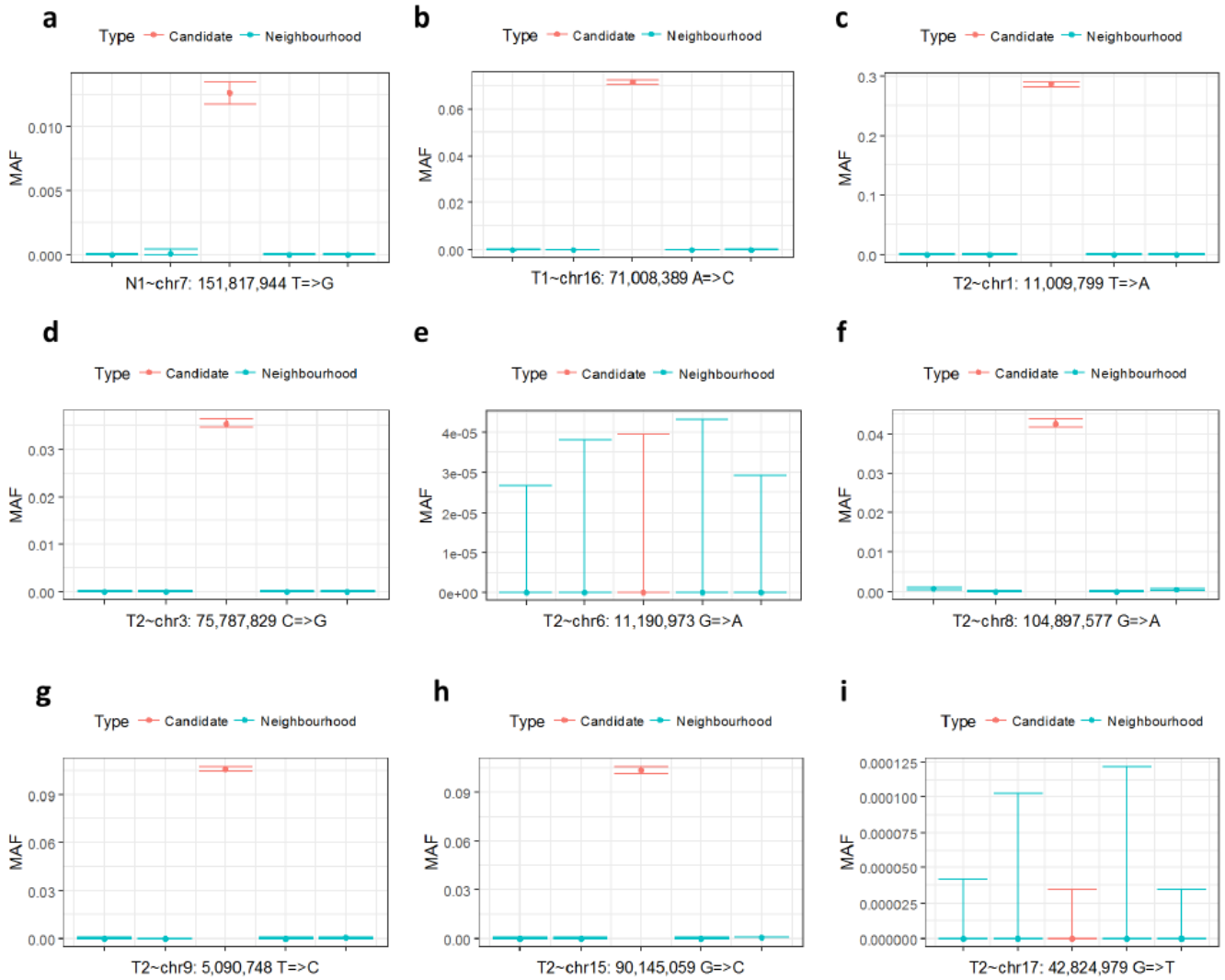


mDDPCR

(Wang, *et al*. *Nat Comms* 2017)

Appendix

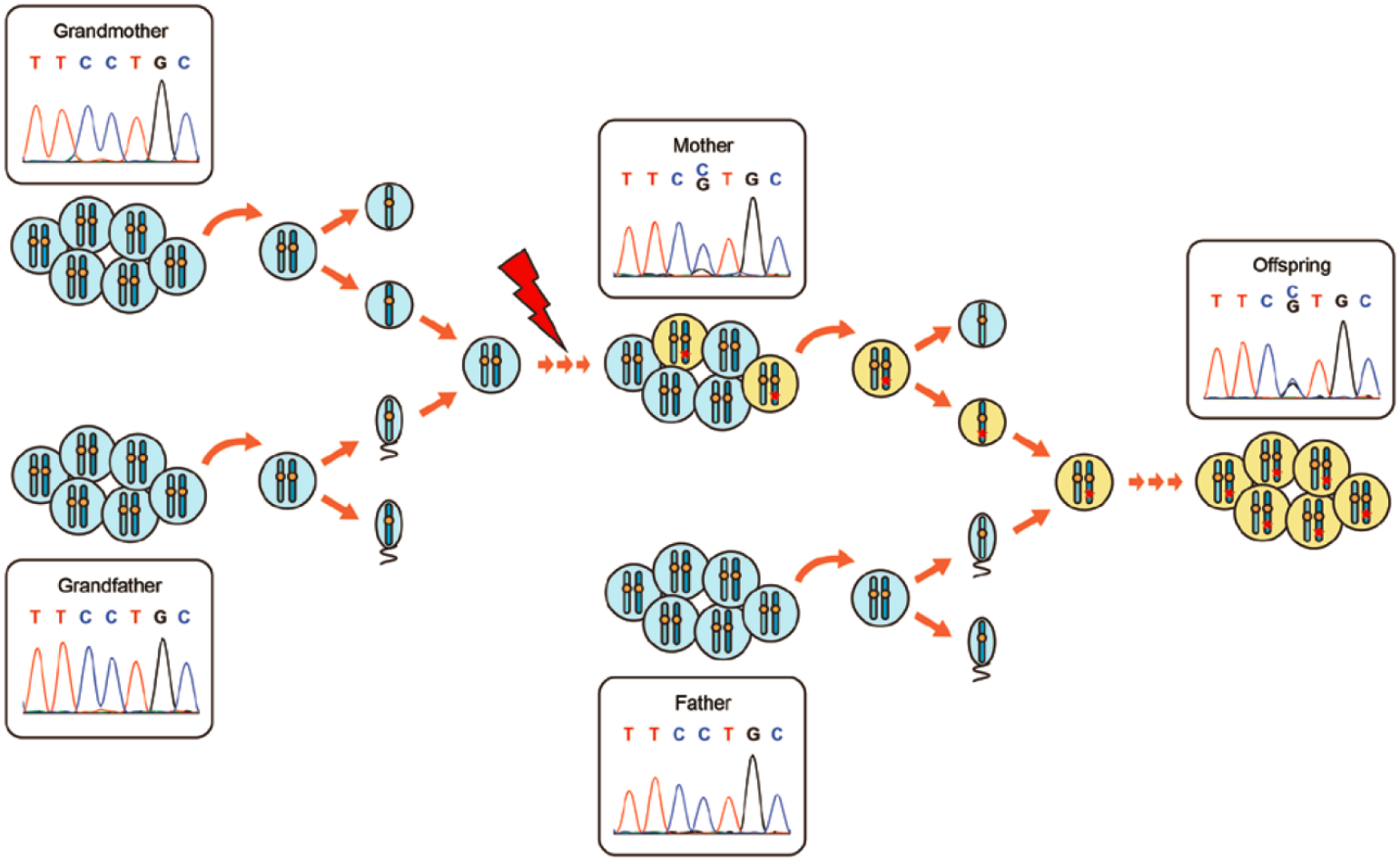
PASM + Ampliseq validated positive and negative low-frequency mutations by o2n-seq



(Wang, *et al*. *Nat Comms* 2017)

Appendix

An example for parental mosaicism transmitted to children and cause severe disorders



(Huang, *et al.*, *Cell Res.*, 2014)