recessive variations [14]. To further explore the missing ASD risk heritability, a promising approach would be to comprehensively identify rare variants that have additive gene effects or show a multigenic epistatic contribution.

Here we have developed a rapid, cost-effective and comprehensive analysis workflow for detecting rare variants in ASD patients. We screened 62 known ASD associated genes using microdroplet PCR-based technology, together with the Ion Torrent Personal Genome Machine (PGM) and MiSeq platforms. To validate the systems, we sequenced 10 positive controls with other diseases and 28 ASD patients. Sequencing data produced by the two sequencers were compared, demonstrating successful identification of positive control variants and novel SNVs associated with ASD.

Materials and Methods

Ethics statement

Written informed consents were obtained from all patients or their parents. Experimental protocols were approved by the Committee for Ethical Issue at Yokohama City University School of Medicine.

Patients

A total of 28 ASD patients, diagnosed according to DSM IV-TR criteria, and 10 patients with other identified diseases with known mutations in one of the target genes, were used for this study. DNA was obtained from peripheral blood leukocytes.

RainDance library preparation and DNA enrichment

The RainDance ASDSeq™ Research Screening Panel was provided by RainDance Technologies™ (Lexington, MA, USA). The RainDance ASDSeq™ panel is a genetic screening tool that offers >92% coverage of 62 genes containing known mutations associated with ASD. The library contains 2349 amplicons ranging in size from 167 to 600 bp and covering a 1034 kb region. Coverage includes all exons for each gene plus 50 bp up- and downstream of each exon, to capture intron/exon splice junctions, as well as 1 kb of both the 5′ promoter region and 3′ UTRs. The panel includes both autosomal and X-linked genes.

A total of 2.5 μg of genomic DNA was used for DNA enrichment. The primer library and a template mix, including 1.5 μg of fragmented genomic DNA and all the PCR reaction components except the primers, were loaded on the RainDance for PCR droplet preparation, according to the manufacturer's instructions. Samples were run on the RDT 1000 machine and PCR droplets were generated. The PCR droplets were amplified under the following conditions: 94°C for 2 min, then 54 cycles of 94°C for 30 sec, 54°C for 30 sec and 68°C for 60 sec, followed by 68°C for 10 min and 4°C for holding. After amplification, the PCR droplets were broken to release the amplicons. The amplicons were purified and quantified using the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The ends of the DNA fragments were repaired at 25°C for 30 min using New England BioLabs End Repair Module (New England BioLabs, Ipswich, MA, USA), followed by purification using Qiagen MinElute columns (Qiagen, Valencia, CA, USA). The PCR fragments were concatenated at 20°C for 30 min using NEB Quick Ligation Kit (New England BioLabs). The ligated products were purified using the Qiagen MinElute columns and fragmented using a Covaris S2 machine (duty cycle 10%, intensity 5, cycle/burst 200, total time per treatment 430 s).

Sequencing using ion torrent PGM and data processing

Library preparation was carried out using the Ion Plus Fragment Library Kit, with 50 ng of amplicons. Adapter ligation, nick repair and amplification were performed as described in the Ion Torrent protocol (Ion Plus Fragment Library Kit; Part Number 4471989 Rev. B; Life Technologies, Grand Island, NY, USA). The Agilent 2100 Bioanalyzer (Agilent Technologies) and associated High Sensitivity DNA kit (Agilent Technologies) were used to determine quality and concentration of the libraries. Emulsion PCR and enrichment steps were carried out using the Ion OneTouch™ Template Kit (Life Technologies) and associated protocol (Part Number 4472430 Rev. C). Sequencing of the amplicon libraries was carried out on the Ion Torrent PGM system using 316 or 318 chips, and barcoding with Ion Xpress™ Barcode Adapters 1–16 Kit (Life Technologies). The Ion Sequencing Kit v2 (Life Technologies) was used for all sequencing reactions (expected read length was 100 bp), following the recommended protocol (Part Number 4469714 Rev. B). After sequencing, reads were mapped to hg19 using Torrent Mapping Alignment Program (TMAP). TMAP is a customized mapping tools for sequencing data generated by PGM, ignoring the indel calls around homopolymer stretch to reduce the hundreds of false negative calls. Torrent Suite 2.0 and/or 3.2 were used for all analyses. Coverage depth was calculated using Torrent Coverage Analysis. SNVs and small insertions/deletions (indels) were identified using the Torrent Variant Caller. Common variants (MAF ≥1%) registered in dbSNP135 (http://www.ncbi.nlm.nih.gov/projects/SNP/) without a flag as clinically associated, or ones in the lower versions of dbSNP, were filtered out. Filter-passed variants were annotated using ANNOVAR [15] and a custom pipeline. In order to compare the ability of mutation detection, reads of positive controls were aligned to GRCh37 with Novoalign v3.00 (Novocraft Technologies, Selangor, Malaysia) with the parameters for PGM and Local realignments around indels and base quality score recalibration were performed using the Genome Analysis Toolkit (GATK) v1.5–21 [16]. SNVs and small indels were identified using the GATK UnifiedGenotyper.

Sequencing using MiSeq and data processing

The same amplicons were sequenced on the Illumina MiSeq sequencer, using the SureSelect^{XT} Reagents (Agilent Technologies) protocol, with 50 ng input material. Each multiplex library pool was sequenced on an Illumina MiSeq for 150 cycles from each end, plus a 6 base-index sequence read, using the MiSeq Reagent Kit (Illumina, San Diego, CA, USA). Image analysis and base calling were performed using sequence control software with real-time analysis, and Consensus Assessment of Sequence and Variation (CASAVA) software v1.8 (Illumina). Reads were aligned to GRCh37 with Novoalign v2.08 (Novocraft Technologies), and Local realignments around indels and base quality score recalibration were performed using the GATK v1.5–21 [16]. SNVs and small indels were identified using the GATK UnifiedGenotyper, and filtered according to the Broad Institute's best-practice guidelines v3. Common variants (MAF ≥1%) registered in dbSNP135 (http://www.ncbi.nlm.nih.gov/projects/SNP/) without a flag as clinically associated, or ones in the lower versions of dbSNP, were filtered out. Filter passed variants were annotated using ANNOVAR [15] and a custom pipeline.

Quality validation of sequence reads

For quality comparison, we combined sequencing data from four random samples obtained by either PGM or MiSeq and evaluated the average quality of data from multiple samples. Box plots for base-call quality of combined runs from each sequencer

were generated using fastqc software (Babraham Bioinformatics, Cambridge, UK). To count the number of single nucleotide polymorphisms (SNPs) and short indels in our combined sequencing data, we used samtools mpileup command with the minimum mapping quality assignment option. We excluded calls with either a depth ≤10 or genotype quality ≤30.

Validation of novel variants

PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/), SIFT (http://sift.jcvi.org/www/SIFT_BLink_submit.html), MutationTaster (http://www.mutationtaster.org/) and Genomic Evolutionary Rate Profiling (GERP) [17] were used to evaluate SNVs in terms of sequence conservation, chemical change and likelihood of pathogenicity. The Human Gene Mutation Database (Biobases, Wolfenbuettel, Germany; (https://portal.biobase-international.com/hgmd/pro/start.php) was used for determining if variants were previously reported.

Sanger confirmation of variants detected by next-generation sequencing

Possible pathological variants were confirmed by Sanger sequencing using an ABI 3500×l or ABI 3100 autosequencer (Life Technologies), according to the manufacturer's protocol. Sequencing data was analyzed using sequence analysis software version 5.1.1 (Applied Biosystems, Foster City, CA, USA) and Sequencher 4.10-build 5828 (GeneCodes Corporation, Ann Arbor, MI, USA).

Statistical analysis

All statistical analyses were carried out using SPSS Statistics 19 (IBM, NY, USA). The carrier frequency of each novel SNV was compared between ASD patients and in-house 212 normal Japanese controls using Fisher's exact test. $p<0.05$ was considered statistically significant.

Results

Sequencing yields and targeting efficiency

The targeted NGS panel was designed to amplify all exons of the 62 known ASD associated genes (Table S1). To validate the performance of RainDance sample enrichment and our chosen NGS systems, ten positive controls, each with a mutation in either *NSD1* (c.3958C>T, c.5177C>T, c.5179G>C, c.6499T>C), *MECP2* (c.243_244insC, c.316C>T), *CASK* (c.277_288del), *SCN1A* (c.342_344delinsAGGAGTT, c.4313T>A) or *CDKL5* (c.145G>A) were used. Our workflow strategy is summarized (Table 1). NGS after target enrichment yielded an average of 295.97 (PGM-TMAP), 201.73 (PGM-Novoalign) and 469.42 (MiSeq) Mb of sequence, in which 96.8% (PGM-TMAP), 78.8% (PGM-Novoalign) and 75% of reads were mapped to the genome, and 26.7% (PGM-TMAP), 28.3% (PGM-Novoalign) and 22.7% were mapped to the targeted regions, by PGM and MiSeq, respectively (Table 2). The percentage of mapped bases was greater in PGM-TMAP than in PGM-Novoalign, while the ones in PGM-Novoalign and MiSeq were similar. On-target rate was also similar and generally low in these data. The total coverage of all targeted bases was on average for PGM (TMAP), 93.7% at 10× and 85.9% at 20×, with a mean read depth of 63×, and for MiSeq, 96.8% at 10× and 93.2% at 20×, with a mean read depth of 95× (Table 2). The complete coverage information on the differences between PGM and MiSeq is presented in Table 2. The mean depth of coverage on genes across all samples ranged from 21× for *PTCHD1* to 237× for *NHS*, with an average of 95× by MiSeq. Despite the high mean read depth and target region coverage, several exons including exon 15 of *NIPBL*, exon 43 of *RELN*, exon 2 of *BRAF*, exon 7 of *PTEN*, exon 10 of *SLC6A4*, exon11 of *SHANK3*, exon 43 of *DMD*, exon 8 of *CASK*, exon 36 of *MED12* and exon 2 of *L1CAM*, had no mapped reads from either sequencer. These unmapped regions may be due to sequence complexity, problematic library synthesis necessitating the use of a concatenation step for sample preparation, or unusual GC content of the fragments for the enrichment system. Exon 11 of *SHANK3* has a very high GC content (80%), while exons 2, 43, 15, and 43, of *BRAF*, *DMD*, *NIPBL*, and *RELN*, respectively, have a very low GC content (<35%), and consequently no mapped reads in the NGS data.

Comparison of sequencing quality

The mean base-call quality score obtained from MiSeq was high through entire reads, with a score >30 (Figure S1A, B). The dispersion of scores among reads at specific positions was relatively small. Conversely, the mean base-call quality score obtained from PGM was >25 at the beginning of reads, but gradually decreased to around 20, at approximately base position 100. The dispersion of scores among reads was larger than those obtained using MiSeq. n addition, read lengths produced by each sequencer were different. With MiSeq, all reads had the expected length of 151 bases, whereas with PGM, read lengths were widely distributed from 60 to 150 bp long, although the expected read length was 100 bp (Figure S1C).

Overall, it appeared that the MiSeq output sequences had a higher base-call quality, but it was difficult to compare the scores derived from each sequencer, as PGM and MiSeq adopt different scoring systems for evaluating base-call quality. MiSeq uses Phred [18], while PGM uses a unique Phred-like system consisting of six predictors whose quality values are correlated with the probability of a base miscall. Therefore we compared the mapping quality of each read from both sequencers, as both sequencers adopt the same scoring system for mapping quality [19]. We summed up the total number of reads with a mapping quality >40 and reads <40, and found 94.5% (MiSeq) and 71.2% (PGM) of aligned reads had a mapping score >40 (Figure S1D).

Next we compared the number of indel calls detected by PGM and MiSeq, in the combined data from four individuals randomly chosen (Table S2). With PGM, 9685 SNPs or indels were called, with 5544 indels calls (57.2%). The frequency of indels was calculated as 1.34 per 1 kb per sample. With MiSeq, 3818 SNPs or indels were called, with 395 calls (10.3%) being indels. The frequency of indels was calculated as 0.096 per 1 kb per sample. After filtering the SNP and indel call with a mapping quality >40, and comparing again, 5288 indels out of 7574 total calls (69.8%) were detected with PGM, while 386 indels out of 3553 total calls (10.9%) were detected with MiSeq, leading to an expected frequency of 1.27 indels per 1 kb per sample (PGM) versus 0.093 indels per 1 kb per sample (MiSeq).

Confirmation of variant detection

The ability of PGM and MiSeq to efficiently detect various mutations, including point mutations and small indels, was tested using previously Sanger-confirmed mutations in variant-positive samples (Table 3). The variant-positive samples included all types of variants, including missense, small insertion, small deletion and small indel variants, in the genes *SCN1A*, *NSD1*, *MECP2*, *CDKL5* and *CASK* (Table 3). Some of the insertion and indel variants detected by NGS are shown (Figure S2A, B). All confirmed variants had a coverage of at least 8× reads, and a mutant allele percentage of 33–62% for heterozygous or 83–100% for hemizygous variants (Table 3). The mutation detection rate was

Table 1. Strategy for validation of RainDance sample enrichment and NGS methods.

	PGM	MiSeq
Number of samples	10	10
Sample enrichment	RDT1000*	RDT1000*
Sequence generated	100 bp single-end** (316 chip/318 chip)	150 bp pair-end (Miseq Reagent Kit)
Mapping	TMAP v2.0.1/Novoalign	Novoalign
SNP/indel identification	Variant caller/GATK	GATK
Annotation	ANNOVAR	ANNOVAR

*The sequencing library used was the RainDance ASDSeq™Research Screening Panel.
**PGM provided the protocol for paired-end sequencing in the end of 2011, only for optional.
doi:10.1371/journal.pone.0074167.t001

either 70% (PGM using standard analysis software of TMAP and Variant Caller) or 100% (MiSeq). With PGM, the variant located near the homopolymer could not be detected because of PGM's high frequency of homopolymer sequencing errors [20,21]. When using TSv3.2 for PGM data analysis, one out of four mutations not identified by TSv2.0, were additionally detected. In order to analyze on the same analytical platforms, sequence data of PGM were also processed using Novoalign for mapping and GATK for variant calling. The mutation detection rate differed significantly between platforms (TMAP-Variant Caller and Novoalign-GATK) (Table 3). Respective PGM data, displayed in the Integrative Genomics Viewer (IGV) [22], showed an increase in sequence mismatch patterns at amplicon ends.

Validation of the RainDance ASD panel for detecting novel mutations in ASD patients

RainDance targeted resequencing was obtained on a total of 28 ASD patients, with a mean total sequence length of 273 or 446 Mb, and an average read depth of approximately 65× or 115×, for PGM and MiSeq, respectively (Table 4). After filtering by dbSNP135, a total of 98 (PGM) and 62 (MiSeq) variants were discovered following RainDance target enrichment. Of these, 62 (PGM) and 46 (MiSeq) were nonsynonymous SNVs (Table S3). Under a rare variant hypothesis, variants were filtered to exclude common variants (MAF ≥1%), using the Exome Variant Server from the NHLBI Exome Sequencing Project and an internal dataset of 212 control exomes from the Japanese population. Although c.878C>T (p.S293F) in *SLC6A4* was detected in 4/212 control exomes (MAF = 0.01%), we chose not to remove this SNV, since it has been functionally proven to disrupt serotonin transporter activity [23]. We validated a total of 57 (PGM) and 30 (MiSeq) SNVs. These SNVs were confirmed by Sanger sequencing, with 21 (PGM) and 22 (MiSeq) shown to be true positives (Table S3). In contrast, after filtering to exclude common variants, no indel mutations were detected by either PGM or MiSeq. All 21 SNVs detected by PGM were also detected by MiSeq. We analyzed the ability of each platform to detect variants and found that both platform was able to identify true variants, but PGM produced many false variant calls. The true positive call rates in the entire coding region were 36.8% (PGM) and 73.3% (MiSeq) (Table S3). We inspected each false positive calls in PGM and MiSeq using IGV to evaluate what kind of errors they were. In PGM, 27/36 calls (75%) had low depth, 21/36 calls (58.3%) had calls at respective read end, 14/36 calls (38.8%) were located near homopolymers, and 1/36 calls (2.7%) had PGM specific low quality error. In MiSeq, 5/8 calls (62.5%) had calls at respective read end and 3/8 calls (37.5%) had MiSeq specific errors. (Table S3).

Table 2. Comparison between PGM and MiSeq sequencing performance in 10 positive controls.

	PGM		MiSeq
	TMAP	Novoalign	
Average total number of bases (Mb)	295.97	201.73	469.42
Average read length (base)	116	116	150
% mapped on human genome	96.8%	78.8%	75%
% on target regions	26.7%	28.3%	22.7%
Mean depth of coverage	63	57	95
% of target regions at >10-fold coverage	93.7%	92.1%	96.8%
% of target regions at >20-fold coverage	85.9%	82.0%	93.2%

doi:10.1371/journal.pone.0074167.t002

Candidate rare SNVs associated with ASD

We identified 22 rare SNVs in 28 patients with ASD (Table 5). Clinical features of the patients with these rare SNVs were demonstrated (Table S4). We considered some to be disease causing, as they are the same mutations previously reported in patients with different diseases that accompany autistic features, namely, c.4612G>A (p.V1538I) in *SCN1A*, identified in a patient with Dravet syndrome [24], and c.878C>T (p.S293F) in *SLC6A4*, identified in a patient with serotonin transporter deficiency [23]. The c.7880G>A (p.R2627Q) mutation identified in *CHD7* was not the same mutation, but was found at the same position, as the one detected in a patient with CHARGE syndrome [25]. Of these three patients, parent samples were only available for the patient with the *SLC6A4* mutation, and the mutation was shown to be inherited from a mother with no autistic features.

Eighteen of the identified SNVs were not observed in 212 in-house Japanese control exomes, suggesting they may be strong candidates for ASD associated SNVs. The remaining four SNVs were also observed in control exomes; however, with a lower frequency than patients with ASD, leading to an OR of 1.93–25.32. In particular, c.56C>T (p.A19V) was detected significantly more frequently in patients with ASD than in controls (OR, 25.32; 95% confidence interval (CI), 2.54–252.76). The remaining SNVs did not reach statistical significance, likely due to the limited number of patients analyzed.

Based on web-based prediction software, 72.7% of the detected SNVs (16/22) were deemed pathogenic by either PolyPhen-2

Table 3. Validation of our chosen NGS methods for mutation detection.

					Detected by			Coverage			Mutant allele (%)		
Sample	Sex	Chr	Gene	Mutation	PGM[1)]	PGM[2)]	MiSeq	PGM[1)]	PGM[2)]	MiSeq	PGM[1)]	PGM[2)]	MiSeq
1	F	2	*SCN1A*	c.342_344delinsAGGAGTT	−	−	+	13	n.a.	91	n.a.	n.a.	44
2	F	2	*SCN1A*	c.4313T>A (p.M1438K)	+	+	+	31	42	48	33	31	38
3	M	5	*NSD1*	c.3958C>T (p.R1320X)	+	−	+	34	n.a.	50	62	n.a.	40
4	M	5	*NSD1*	c.5177C>T (p.P1726L)	+	−	+	37	n.a.	93	38	n.a.	46
5	M	5	*NSD1*	c.5179G>C (p.A1725P)	+	−	+	55	n.a.	62	47	n.a.	50
6	M	5	*NSD1*	c.6499T>C (p.C2167R)	+	−	+	77	n.a.	223	46	n.a.	54
7	F	X	*MECP2*	c.243_244insC	−	−	+	18	n.a.	123	n.a.	n.a.	41
8	F	X	*MECP2*	c.316C>T (p.R106W)	+	−	+	60	n.a.	76	42	n.a.	47
9	M	X	*CDKL5*	c.145G>A (p.E49K)	−	−	+	8	n.a.	46	n.a.	n.a.	100
10	M	X	*CASK*	c.227_228del	(+)	+	+	35	47	112	83	81	97

F, Female; M, Male; Chr, Chromosome; +, Detected; −, Not detected; (+), Mutation only detected by TSv3.2, and not by TSv2.0.; n.a., Not applicable;
[1)]Reads were mapped by TMAP and SNVs and indels were identified using the Torrent Variant Caller.
[2)]Reads were mapped by Novoalign v3.00 and SNVs and indels were identified using the GATK v1.5–21.
doi:10.1371/journal.pone.0074167.t003

(36.3%; 8/22 SNVs), SIFT (50%; 11/22 SNVs), or MutationTaster (13.5%; 3/22 SNVs). We annotated positions with their conservation as scored with the GERP. Mutations at highly conserved positions would be predicted to be functionally important (45.5%; 10/22 SNVs).

Five out of 28 patients had multiple SNVs (Table S5). Following the multigenic contribution theory in ASD [4], these could be associated with the onset or the severity of this disease.

Discussion

We have developed an efficient workflow for detecting rare SNVs/indels in ASD associated genes using bench-top next generation sequencers with target gene enrichment. The evaluation and comparison of NGS devices are of recent interest to us. In this study we chose to compare the Ion Torrent PGM and Illumina MiSeq, which are currently the most popular NGS. The characteristics of the two devices are shown (Table S6). In this study, we compared the sequence yield and quality of these two NGS platforms, and showed a practical use for targeted resequencing of human genes.

Our comparison of two bench-top sequencers showed their yields were both greater than expected; however, the quality of sequence reads varied: better than expected through entire reads in MiSeq, while barely exceeding the minimum expected quality value with large discrete reads in PGM. Comparing the mapping quality of the two sequencers, which was calculated based on the same algorithm, the percentage of reads with a mapping quality ≤40 was markedly more in PGM than in MiSeq. Considering their target regions were the same, this difference reflects the difference of overall read quality from the two sequencers. Focusing on indel calls, we found an excess with PGM, compared to MiSeq. The number of MiSeq indel calls is reasonable, compared to the estimated error rate (0.11 to 0.08 per 1 kb) in conventional capillary sequencing of the human genome [26]. Even with filtering of the reads for low genotyping quality and depth, the excess indel calls in PGM did not decrease. As previously reported, excess indel calls or a lower read quality are considered to be largely due to homopolymers [20,27]. This unique characteristics of PGM was reflected in the difference of mapped rates for PGM-generated data when using different mapping tools, TMAP or Novoalign. As shown in Table 2, the mapped rates of bases between PGM-generated data and MiSeq-generated data using Novoalign were similar, being reasonable since these two data were derived from the same sample libraries, while the one for PGM-generated data using TMAP was better. We assume this is because TMAP consider homopolymer-associated indel errors on mapping and could map more reads which standard mapping tools such as Novoalign could not. The difference in the mapped rates for PGM-generated data might affect the mutation detection rate. Based on the difference in mutation detection rates of positive controls in PGM-generated data with different pipelines (Table 3), custom mapping and the SNP/indel detection software, TMAP and Variant Caller, are necessary for the PGM workflow to reduce mapping errors without compromising detection sensitivity. Otherwise the number of false positive indel calls would be greatly increased.

Table 4. Comparison between PGM and MiSeq sequencing performance in 28 ASD patients.

	PGM	MiSeq
Average of total number of bases (Mb)	273.06	445.99
% on target regions	30.20%	25.60%
Mean depth of coverage	65	115
% of target regions at >10-fold coverage	92.70%	95.50%

doi:10.1371/journal.pone.0074167.t004

Generally, target gene enrichment using the RDT machine worked well, but there were some disadvantages, including a relatively low on-target rate as shown in Table 2, and occasional sample enrichment failure. This may be partially due to the genomic complexity or a biased GC content of target regions. Alternatively, it may be due to the screening panel itself, which does not employ a tailed primer system using PCR amplification primers, therefore necessitating the use of the concatenation step for sample preparation.

In our workflow validation using ten positive controls, the mutation detection rate was lower with PGM than MiSeq. False

Table 5. Rare SNVs identified with amino acid changes and computational predictions of pathogenicity.

Gene	Accession No.	Nucleotide : amino acid change	MutationTaster	Polyphen2[1] (Hum Div)	SIFT[2]	GERP[3]	HGMD[4]	genotype (allele) cases	genotype (allele) controls	OR (95% CI)	*p* value	Patient
BRAF	NM_004333	c.976A>G:p.I326V	polymorphism	0	0.71	−5.32	none	1/28 (1/56)	1/212 (1/424)	7.82 (0.46–128.60)	0.22	A682
CACNA1C	NM_001129837	c.4706C>T:p.P1569L	polymorphism	0.001	**0.04**	2.39	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A681
CHD7	NM_017780	c.7880G>A:p.R2627Q	polymorphism	**0.997**	**0.01**	**5.56**	CHARGE syndrome (R2627X)	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A634
CHD7	NM_017780	c.7652C>A:p.T2551N	polymorphism	0.01	0.31	**5.63**	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A447
CNTNAP2	NM_014141	c.1276C>A:p.L426I	polymorphism	**0.977**	**0.03**	**5.7**	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A479
CNTNAP2	NM_014141	c.1448G>A:p.R483Q	polymorphism	**0.991**	0.4	**5.07**	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A621
DMD	NM_004007	c.3479A>G:p.N1160S	polymorphism	**0.973**	0.22	1.36	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A668
DMD	NM_004007	c.2473A>G:p.M825V	polymorphism	0.026	0.36	3.87	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A668
MID1	NM_001193278	c.555G>A:p.M185I	polymorphism	**0.839**	**0.01**	**5.64**	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A669
NIPBL	NM_015384	c.1553C>T:p.T518I	polymorphism	0.275	**0**	**5.88**	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A681
NRXN1	NM_001135659	c.455G>A:p.G152D	**disease causing**	0	1	4.97	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A711
NSD1	NM_022455	c.2087T>C:p.V696A	polymorphism	0.189	**0.02**	3.94	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A464
PNKP	NM_007254	c.56C>T:p.A19V	polymorphism	0.026	0.27	4.55	none	3/28 (3/56)	1/212 (1/424)	**25.32 (2.54–252.76)**	**0.005**	A627, A651, A674
RAI1	NM_030665	c.1148C>T:p.P383L	polymorphism	1	**0**	**5.55**	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A663
RAI1	NM_030665	c.4238T>C:p.M1413T	polymorphism	0.011	**0**	2.93	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A634
RELN	NM_005045	c.8915A>C:p.K2972T	**disease causing**	**0.996**	**0.03**	**5.89**	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A653
SCN1A	NM_001165963	c.4612G>A:p.V1538I	polymorphism	**0.89**	0.09	**5.76**	Dravet syndrome	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A695
SHANK3	NM_033517	c.3169C>T:p.L1057F	polymorphism	0.232	0.27	3.19	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A668
SLC6A4	NM_001045	c.878C>T:p.S293F	**disease causing**	0	0.25	3	Serotonin transporter deficency	1/28 (1/56)	4/212 (4/424)	1.93 (0.21–17.87)	0.47	A674
TSC2	NM_000548	c.2032G>A:p.A678T	polymorphism	0.016	0.23	−0.706	none	1/28 (1/56)	1/212 (1/424)	7.82 (0.48–128.60)	0.22	A647
VPS13B	NM_015243	c.820T>G:p.F274V	polymorphism	0.314	**0**	**5.45**	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A663
VPS13B	NM_017890	c.11960C>G:p.P3987R	polymorphism	0.437	0.06	3.85	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A619

[1]PolyPhen2 scores close to 1 are likely to be pathogenic (highlighted in bold). HumDiv-trained Polyphen-2 assumes even mildly deleterious alleles as damaging to evaluate rare alleles potentially involved in complex phenotypes.
[2]SIFT scores less than 0.05 are likely to be pathogenic (highlighted in bold).
[3]GERP scores above 5 are highly conserved (highlighted in bold).
[4]The Human Gene Mutation Database (HGMD) was searched to identify SNVs registered as disease causing mutations. Carrier frequencies of each SNV were statistically compared between ASD patients (cases) and in-house normal 212 controls (controls). Results are presented as odds ratio's (OR) and p values. Pathogenic findings are shown in bold. CI, confidence interval; wt, wild type allele; mut, mutant allele; n.d., not determined.
doi:10.1371/journal.pone.0074167.t005

negatives are largely due to the weakness in indel detection, implying not only excess false positive, but also increased false negative indel calls with PGM. Another typical false negative mutation identified with PGM was detected at amplicon ends. This may happen more readily with PGM as the read length is not as long as expected. On the other side, the higher coverage of the MiSeq data is expected due to the longer read lengths as well as paired end reads. With regards to SNV detection, both PGM and MiSeq showed high mutation detection rates (6/7 mutations, 85.7% in PGM vs. 7/7 mutations, 100% in MiSeq). Target resequencing of 28 patients with ASD identified 21 candidate SNVs in PGM versus 22 in MiSeq, again showing similar SNV variant detection abilities. Although there is a higher false-positive SNV call rate with PGM compared to MiSeq due to the same factors observed in positive control studies, At present it would be reasonable to apply PGM for SNV detection. Recent rapid updates of the device, chemistry and mapping/mutation detection software in PGM may potentially reduce these drawbacks in the near future.

ASD is a genetically heterogeneous disease, with a complex genetic architecture [4]. In particular, rare SNVs with a multigenic contribution are expected to play a specific role in the molecular pathogenesis of ASD. We have shown that our workflow works rapidly and inexpensively to address this issue by demonstrating our successful identification of novel candidate SNVs in ASD. Notably, A19V in *PNKP* was identified significantly more in patients with ASD than controls. PNKP (polynucleotide kinase 3′-phosphatase) is a bi-functional enzyme that possesses both DNA 3′-phosphatase and DNA 5′-kinase activities, and associates with the single strand break repair machinery. Single strand break could be hazardous to the cell if left unrepaired, especially in central nervous system since frequently single strand breaks could happen [28]. *PNKP* is mutated in microcephaly, early-onset, intractable seizures and developmental delay (MCSZ), in autosomal recessive manner. Patients with MCSZ sometimes show variable behavioral problems, mainly hyperactivity [29]. Considering enzymatic activity of PNKP and its stability as reported [30], clinical symptoms of individuals with the heterozygous variant may not be as severe as MCSZ, however it could not be denied that possible decrease in enzyme activity or protein level of PNKP comparing to wild type might affect the normal development of central nervous system. It was implied that *PNKP* might be a candidate for ASD-related gene by copy number analysis previously [31]. We showed for the first time a candidate variant associated with ASD. Further study with larger samples is necessary to confirm its pathogenicity. It is also noted that there were some genes such as *CHD7*, *CNTNAP2*, *DMD*, and *RAI1*, in which two patients had private rare variants. It is speculated that the private variants of those might accumulate in ASD populations.

In conclusion, we present the comparison of two bench-top sequencers, PGM and MiSeq, through the newly developed workflow for the investigation of ASD. Analyzing larger sample sets may lead to unraveling of the missing heritability of ASD.

Supporting Information

Figure S1 Comparison of overall sequencing quality between PGM and MiSeq. (A) Box plots of base-call quality scores across all bases obtained using PGM with a 316 chip (left panel) or MiSeq (right panel). Green and red areas indicate quality scores above 28 and below 20, respectively. Yellow boxes show upper and lower quartiles with whiskers indicating 10% and 90% quartiles. Red horizontal lines indicate the median value. Blue curves represent the mean quality scores. Quality scores are given based on the calculation of Phred-scaled quality values using $q = -10\log10(P)$, with P being the estimated error probability for that base-call. (B) Quality score distribution over all sequence reads obtained using PGM with a 316 chip (left panel in red) or MiSeq (right panel in blue). Combined data from four samples are displayed. Mean quality scores across all base-calls from a particular sequence, calculated as the Phred score, are shown on the X axis, and the number of reads with the specified mean sequence quality on the Y axis. (C) Distribution of read length from all sequence reads obtained using PGM with a 316 chip (left panel in red) or MiSeq (right panel in blue). Read lengths are shown on the X axis, and the number of reads with the specified read lengths on the Y axis. (D) Mapping quality from all sequence reads obtained using PGM with a 316 chip (red bars) or MiSeq (blue bars). The number of reads with a mapping quality of either <40 or ≥ 40 in each device (left panel). The percentage of reads with mapping quality ≥ 40 in each device (right panel). MQ, mapping quality.
(TIF)

Figure S2 Comparison between PGM and MiSeq of mutations and sequence reads from positive control samples. (A) The c.342_344delinsAGGAGTT mutation detected in Sample 1. (B) The c.243_244insC mutation detected in Sample 7. In both panels, data was obtained from either PGM (upper) or MiSeq (lower). Both the c.342_344delinsAGGAGTT mutation and the c.243_244insC mutation were not detected in PGM with neither PGM-TMAP-Variant Caller algorithm nor PGM-Novolign-GATK algorithm. Forward and reverse read strands are shown in pink and blue, respectively. Red and blue arrows indicate insertion and deletion positions, respectively, which were confirmed by Sanger sequencing. The horizontal bar indicates the deletion call, and symbols within the read strands (I·) indicate insertion calls detected by either PGM or MiSeq. In (A) and (B), the true inserted sequence depicted by "I·" commonly detected by PGM and MiSeq is AACTCC and C, respectively. The DNA sequence surrounding a mutation is shown below the IGV graphics. WT, wild type; Pt, patient.
(TIF)

Table S1 RainDance ASDSeq™ Core Research Screening Panel.
(PDF)

Table S2 Summary of SNP/indel detection with PGM and MiSeq.
(PDF)

Table S3 Summary of target resequencing and prioritization.
(PDF)

Table S4 Clinical features of patients with novel SNVs.
(PDF)

Table S5 Multiple mutations detected in patients with ASD.
(PDF)

Table S6 Comparison of PGM and MiSeq analysis cost and expected yield.
(PDF)

Acknowledgments

We thank all the participants for their cooperation in this research. We also thank Ms. Y. Yamashita, Ms. K. Takabe and Mr. T. Miyama from the

Department of Human Genetics, Yokohama City University Graduate School of Medicine, and Dr. D. Yamaguchi from BITS, for their technical assistance.

Author Contributions

Conceived and designed the experiments: EK SM N. Matsumoto. Performed the experiments: EK SM. Analyzed the data: EK SM MN YT N. Miyake HS. Contributed reagents/materials/analysis tools: NO. Wrote the paper: EK SM N. Matsumoto.

References

1. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, et al. (2011) Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 70: 863–885.
2. Autism and Developmental Disabilities Monitoring Network Surveillance Year 2008 Principal Investigators CDC (2012) Prevalence of autism spectrum disorders-Autism and Developmental Disabilities Monitoring Network. MMWR Surveill Summ 61: 1–19.
3. American Psychiatric Association (2000) Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition-Text Revision (DSMIV-TR); Association AP, editor. Washington D.C.
4. Devlin B, Scherer SW (2012) Genetic architecture in autism spectrum disorder. Curr Opin Genet Dev 22: 229–237.
5. Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, et al. (2011) Genetic heritability and shared environmental factors among twin pairs with autism. Arch Gen Psychiatry 68: 1095–1102.
6. Stein JL, Parikshak NN, Geschwind DH (2013) Rare inherited variation in autism: beginning to see the forest and a few trees. Neuron 77: 209–211.
7. Klei L, Sanders SJ, Murtha MT, Hus V, Lowe JK, et al. (2012) Common genetic variants, acting additively, are a major source of risk for autism. Mol Autism 3: 9.
8. Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, et al. (1995) Autism as a strongly genetic disorder: evidence from a British twin study. Psychol Med 25: 63–77.
9. Anney R, Klei L, Pinto D, Almeida J, Bacchelli E, et al. (2012) Individual common variants exert weak effects on the risk for autism spectrum disorderspi. Hum Mol Genet 21: 4781–4792.
10. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, et al. (2012) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485: 237–241.
11. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, et al. (2012) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485: 246–250.
12. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, et al. (2012) Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 485: 242–245.
13. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, et al. (2012) De novo gene disruptions in children on the autistic spectrum. Neuron 74: 285–299.
14. Lim ET, Raychaudhuri S, Sanders SJ, Stevens C, Sabo A, et al. (2013) Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. Neuron 77: 235–242.
15. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38: e164.
16. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43: 491–498.
17. Cooper GM, Goode DL, Ng SB, Sidow A, Bamshad MJ, et al. (2010) Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. Nat Methods 7: 250–251.
18. Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8: 186–194.
19. Li H, Ruan J, Durbin R (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res 18: 1851–1858.
20. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, et al. (2012) Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 30: 434–439.
21. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, et al. (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13: 341.
22. Thorvaldsdottir H, Robinson JT, Mesirov JP (2012) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform.
23. Prasad HC, Zhu CB, McCauley JL, Samuvel DJ, Ramamoorthy S, et al. (2005) Human serotonin transporter variants display altered sensitivity to protein kinase G and p38 mitogen-activated protein kinase. Proc Natl Acad Sci U S A 102: 11545–11550.
24. Depienne C, Trouillard O, Saint-Martin C, Gourfinkel-An I, Bouteiller D, et al. (2009) Spectrum of SCN1A gene mutations associated with Dravet syndrome: analysis of 333 patients. J Med Genet 46: 183–191.
25. Jongmans MC, Admiraal RJ, van der Donk KP, Vissers LE, Baas AF, et al. (2006) CHARGE syndrome: the phenotypic spectrum of mutations in the CHD7 gene. J Med Genet 43: 306–314.
26. Meader S, Hillier LW, Locke D, Ponting CP, Lunter G (2010) Genome assembly quality: assessment and improvement using the neutral indel model. Genome Res 20: 675–684.
27. Liu L, Li Y, Li S, Hu N, He Y, et al. (2012) Comparison of next-generation sequencing systems. J Biomed Biotechnol 2012: 251364.
28. Reynolds JJ, Stewart GS (2013) A single strand that links multiple neuropathologies in human disease. Brain 136: 14–27.
29. Shen J, Gilmore EC, Marshall CA, Haddadin M, Reynolds JJ, et al. (2010) Mutations in PNKP cause microcephaly, seizures and defects in DNA repair. Nat Genet 42: 245–249.
30. Reynolds JJ, Walker AK, Gilmore EC, Walsh CA, Caldecott KW (2012) Impact of PNKP mutations associated with microcephaly, seizures and developmental delay on enzyme activity and DNA strand break repair. Nucleic Acids Res 40: 6608–6619.
31. Gregory SG, Connelly JJ, Towers AJ, Johnson J, Biscocho D, et al. (2009) Genomic and epigenetic evidence for oxytocin receptor deficiency in autism. BMC Med 7: 62.

Whole genome sequencing in patients with retinitis pigmentosa reveals pathogenic DNA structural changes and *NEK2* as a new disease gene

Koji M. Nishiguchi[a,b], Richard G. Tearle[c], Yangfan P. Liu[d], Edwin C. Oh[d,e], Noriko Miyake[f], Paola Benaglio[a], Shyana Harper[g], Hanna Koskiniemi-Kuendig[a], Giulia Venturini[a], Dror Sharon[h], Robert K. Koenekoop[i], Makoto Nakamura[b], Mineo Kondo[b], Shinji Ueno[b], Tetsuhiro R. Yasuma[b], Jacques S. Beckmann[a,j,k], Shiro Ikegawa[l], Naomichi Matsumoto[f], Hiroko Terasaki[b], Eliot L. Berson[g], Nicholas Katsanis[d], and Carlo Rivolta[a,1]

[a]Department of Medical Genetics, University of Lausanne, 1005 Lausanne, Switzerland; [b]Department of Ophthalmology, Nagoya University School of Medicine, Nagoya 466-8550, Japan; [c]Complete Genomics, Inc., Mountain View, CA 94043; [d]Center for Human Disease Modeling and [e]Department of Neurology, Duke University, Durham, NC 27710; [f]Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama 236-0004, Japan; [g]Berman-Gund Laboratory for the Study of Retinal Degenerations, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Boston, MA 02114; [h]Department of Ophthalmology, Hadassah-Hebrew University Medical Center, Jerusalem 91120, Israel; [i]McGill Ocular Genetics Laboratory, McGill University Health Centre, Montreal, QC, Canada H3H 1P3; [j]Service of Medical Genetics, Lausanne University Hospital, 1011 Lausanne, Switzerland; [k]Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland; and [l]Laboratory for Bone and Joint Diseases, Center for Genomic Medicine, RIKEN, Tokyo 108-8639, Japan

Edited by Jeremy Nathans, Johns Hopkins University, Baltimore, MD, and approved August 15, 2013 (received for review May 1, 2013)

We performed whole genome sequencing in 16 unrelated patients with autosomal recessive retinitis pigmentosa (ARRP), a disease characterized by progressive retinal degeneration and caused by mutations in over 50 genes, in search of pathogenic DNA variants. Eight patients were from North America, whereas eight were Japanese, a population for which ARRP seems to have different genetic drivers. Using a specific workflow, we assessed both the coding and noncoding regions of the human genome, including the evaluation of highly polymorphic SNPs, structural and copy number variations, as well as 69 control genomes sequenced by the same procedures. We detected homozygous or compound heterozygous mutations in 7 genes associated with ARRP (*USH2A*, *RDH12*, *CNGB1*, *EYS*, *PDE6B*, *DFNB31*, and *CERKL*) in eight patients, three Japanese and five Americans. Fourteen of the 16 mutant alleles identified were previously unknown. Among these, there was a 2.3-kb deletion in *USH2A* and an inverted duplication of ~446 kb in *EYS*, which would have likely escaped conventional screening techniques or exome sequencing. Moreover, in another Japanese patient, we identified a homozygous frameshift (p.L206fs), absent in more than 2,500 chromosomes from ethnically matched controls, in the ciliary gene *NEK2*, encoding a serine/threonine-protein kinase. Inactivation of this gene in zebrafish induced retinal photoreceptor defects that were rescued by human *NEK2* mRNA. In addition to identifying a previously undescribed ARRP gene, our study highlights the importance of rare structural DNA variations in Mendelian diseases and advocates the need for screening approaches that transcend the analysis of the coding sequences of the human genome.

medical genetics | ophthalmology | ciliopathy | retinal blindness

The identification of the genetic causes of rare Mendelian diseases is becoming increasingly important following some success with gene-based therapy, as recently reported for patients with a form of Leber congenital amaurosis (LCA), a severe autosomal recessive hereditary retinal dystrophy (1–3). The evidence that restoring a gene in the diseased retina could yield therapeutic effects has stimulated the pursuit of the genetic causes of other retinal dystrophies, including retinitis pigmentosa (RP).

RP is the name given to a group of hereditary retinal conditions in which degeneration of rod photoreceptors, responsible for vision under starlight or moonlight conditions, is more pronounced than that of cone photoreceptors, which mediate daylight vision. Individuals with RP typically experience night blindness at first, followed by progressive and unstoppable visual impairment in daytime conditions as well (4). Their visual fields become reduced gradually and sight is lost from the midperiphery to the periphery and then from the midperiphery to the center, resulting eventually in complete or near-complete blindness if left untreated. Most patients show intraretinal pigment in a bone spicule configuration around the fundus periphery, for which this condition was named. In addition, they typically show retinal arteriolar attenuation, elevated final dark adapted thresholds, and reduced and delayed electroretinograms (ERGs) (4). Vitamin A supplementation in combination with an omega-3 rich diet can slow the course of retinal degeneration and preserve visual acuity among adults with this condition (5, 6). Autosomal, recessively inherited RP (ARRP) is the most common form of hereditary retinal degeneration in humans. To date, over 50 genes have been associated with ARRP and allied disorders, among patients who are predominantly of European ancestry (RetNet; www.sph.uth.tmc.edu/retnet/home.htm). However, despite this high number of identified disease genes, ~40–50% of all diagnosed cases have no mutations in recognized loci (7). Furthermore, genetic defects in RP are also population specific. For example, a screening of 193 unrelated Japanese patients with isolate or autosomal recessive RP

Significance

Retinitis pigmentosa (RP) is a genetic disease that causes progressive blindness and that is caused by mutations in more than 50 genes. Conventional methods for identification of both RP mutations and novel RP genes involve the screening of DNA sequences spanning coding exons. In our work, we conversely test the use of whole genome sequencing, a technique that takes into account all variants from both the coding and noncoding regions of the human genome. In our approach, we identify a number of unique RP mutations, a previously undescribed disease gene, as well as pathogenic structural DNA rearrangements originating in introns.

Author contributions: K.M.N. and C.R. designed research; K.M.N., Y.P.L., E.C.O., N. Miyake, P.B., H.K.-K., and G.V. performed research; S.H., D.S., R.K.K., M.N., M.K., S.U., T.R.Y., S.I., N. Matsumoto, H.T., and E.L.B. contributed new reagents/analytic tools; K.M.N., R.G.T., Y.P.L., E.C.O., N. Miyake, P.B., H.K.-K., G.V., J.S.B., S.I., N. Matsumoto, N.K., and C.R. analyzed data; and K.M.N., E.C.O., J.S.B., E.L.B., N.K., and C.R. wrote the paper.

Conflict of interest statement: R.G.T. is an employee and shareholder of Complete Genomics, Inc.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: carlo.rivolta@unil.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1308243110/-/DCSupplemental.

MEDICAL SCIENCES

for 30 disease genes identified commonly within North American or European patients revealed candidate pathogenic mutations in only 14% of the cohort (8).

Recent advances in massively parallel sequencing have enabled the analysis of large amounts of sequences (genes) at reasonable costs, revolutionizing the traditional approach of exon-by-exon Sanger sequencing (9). The two major forms of sequencing strategies allowing large-scale analyses are whole genome sequencing (WGS) and whole exome sequencing (WES). The former reads the entire genome with no distinction between exons and non-exonic regions. It allows the detection of intergenic variants, copy number variations (CNVs), and other structural rearrangements, as well as unrecognized exonic sequences. The latter technique relies on targeted DNA capture and focuses on the analysis of the known exonic content of the genome, performed according to the genomic annotation available at a given point in time.

In this study, we performed WGS as a method for mutation discovery in a highly genetically heterogeneous Mendelian disease; to this end, we evaluated 16 unrelated RP patients from diverse ethnic backgrounds.

Results

Genome Sequencing. Genome sequencing in the 16 analyzed patients produced an average mapping yield of 200.8 ± 17.9 (mean ± SD) Gb and an average coverage of 66.1 ± 2.4 (mean ± SD) reads per base (*SI Appendix*, Table S1). This covered a genomic fraction of 0.968 ± 0.004, in which roughly 3.8 million putative variations were identified. Of these, ~7.7% were not reported in dbSNP build 131 and were classified as novel variants. Variations present within transcripts were classified further as synonymous and nonsynonymous, and analyzed separately for the North American and Japanese sets of patients. Scoring of large structural variations (SVs) could be achieved only for seven genomes, as the remaining DNA samples, possibly because of their older age, did not produce reliable mate pair information (*SI Appendix*, *Results S1*).

Assessment of pathogenic variants was performed by a series of filtering steps, summarized in Fig. 1.

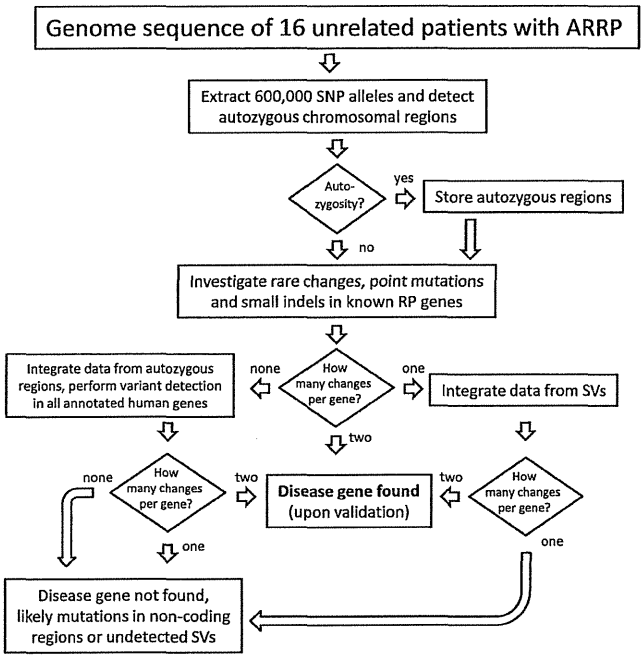

Fig. 1. Flowchart of the filtering process applied in this work.

Assessment of Autozygous Regions. Each genome was evaluated for known or undocumented parental consanguinity as well as for possible founder mutation events by extracting genotypes of known polymorphic SNPs and by searching for long intervals with high degrees of homozygosity (at least 500 consecutive SNP markers, or ~2.2 Mb on average), indicative of identity by descent (IBD). Significant genomic homozygosity was observed only in the five Japanese patients (individual IDs: R14, R15, R16, R18, and R19) who had documented parental consanguinity. The areas of IBD had essentially no overlap among these patients except for a 10-Mb interval on chromosome 1 shared by R15 and R19. Haplotype analysis indicated the shared intervals to be of different origins. No other patients carried genomic areas indicative of IBD; this was consistent with their family history reporting no parental consanguinity.

Sequence Analyses of Known RP Genes. We first focused our analyses on genes known to be associated with ARRP. We investigated both small variants (from 1 to 50 bp) from the mapping of short reads and, whenever possible, large SVs. Our results are summarized in *SI Appendix*, Table S2; detailed results are provided in *SI Appendix*, *Results S1*, Figs. S1 and S3, and Table S3.

In addition to point mutations and short indels (insertion/deletions), we detected pathogenic SVs in *USH2A* and *EYS* in patients 003–019 and R9, respectively, by combining information from sequence coverage and abnormal junctions/mate pair distance. In the genome of patient 003–019, we identified a ~2-kb deletion that removed exon 27 of *USH2A*, whereas patient R9 was found to carry a 446-kb head-to-head inverted duplication of the portion of chromosome 6 that included exons 23–29 of *EYS* (Fig. 2).

We found two pathogenic alleles, in either a homozygous or compound heterozygous state, in 8 of the 16 patients, 5 Americans and 3 Japanese, in seven different genes (*SI Appendix*, Table S2). Six patients carried mutations in one of the following genes: *USH2A*, *RDH12*, *CNGB1*, *EYS*, *PDE6B*, and *DFNB31*; 2 patients had mutations in *CERKL*. None of these mutations were found in the control cohorts of 95 healthy North American or 95 Japanese individuals. None of these mutations were reported previously, except p.R257X in *CERKL* and p.G76R in *RDH12* (10, 11). All mutations cosegregated with RP as recessive, pathogenic alleles in all family members of the index patients for whom DNA samples were available (Fig. 3).

Systematic Screening of All Genes. Based on the data from the analysis of known RP genes, we adopted a pipeline to perform a systematic analysis targeting all annotated genes in the genomes of patients with unsolved genetic etiology (*SI Appendix*, Fig. S2). With the aim of selecting a restricted number of candidate genes, more aggressive filtering was adopted with respect to the one used for the screening of known disease genes. The major differences in the analytical pipeline included removal of all entries in dbSNP. We safely applied this filtering because, given the low frequency of individual mutations in ARRP genes (including undetected ones), the risk of eliminating pathogenic DNA variants that could be fortuitously included in dbSNP build 131 is negligible. Further, to validate this approach, we applied it again retrospectively to the genomes for which mutations in RP genes were already detected. All of identified RP mutations were present in the final list of variants, supporting the sensitivity of the strategy. Detailed results are provided in *SI Appendix*, *Results S2* and are summarized in *SI Appendix*, Figs. S2 and S3 and Table S4.

In R19, in whom we did not find any clear-cut mutations in known ARRP genes, we found a homozygous frameshift variant (p.L206fs, c.617_624delTGTATGAGinsA) in the never in mitosis gene A (NIMA)-related kinase 2 (*NEK2*) gene. This variant was present within a highly homologous genomic stretch of 19.6 Mb of chromosome 1q32, predicted to be IBD (*SI Appendix*, Fig. S4).

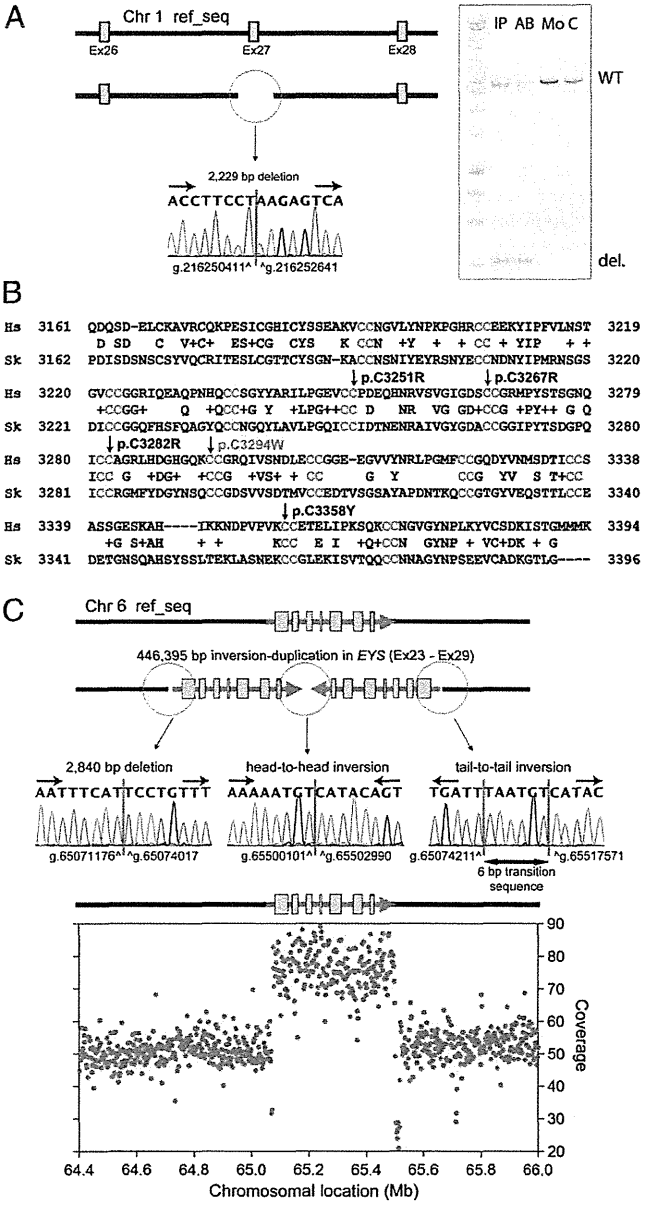

Fig. 2. Pathogenic structural variations identified. (*A*) Sequence of the heterozygous *USH2A* 2,229-bp deletion in patient 003–019 (*Left*) and electrophoresis of the PCR fragments showing a smaller fragment carrying the deletion in the index patient (IP) and her affected brother (AB) but not in her mother (Mo) or a control DNA (C). del, deleted; WT, wild type. (*B*) Alignment of the USH2A protein from *Homo sapiens* (Hs) and *Saccoglossus kowalevskii* (Sk, acorn worm) showing the conservation of 13 CC repeat motifs (red) and the location of the mutation p.C3294W, newly identified in patient 003–019 and her sister. Four previously reported disease-associated missense changes (p.C3251R, p.C3267R, p.C3282R, and p.C3358Y) also affect neighboring CC repeats. (*C*) Schematic representation and DNA sequence of the junctions characterizing the chromosomal rearrangement detected in patient R9 and involving the *EYS* gene. Integration of the information obtained by Sanger DNA sequencing and WGS coverage of the region allows identifying an inverted duplication encompassing exons 23–29.

Similar to most frameshifts producing a premature termination codon, p.L206fs is predicted to result in an mRNA allele that is subject to nonsense-mediated mRNA decay, and therefore in no protein product. Targeted DNA screening revealed that c.617_624delTGTATGAGinsA was absent from 1,273 Japanese and 95 North American control individuals. The entire coding sequence of the *NEK2* gene was then analyzed in a mixed cohort of 190 American patients with ARRP, in 64 Japanese patients with isolate RP, as well as in 13 patients found previously to show linkage between recessive retinal degeneration and the *NEK2* region. However, other than known polymorphisms (rs1056729, rs12031285, and rs45623136), we found only a few isolated heterozygous missense variants (p.R26Q, c.77G>A; p.V137I, c.409G>A; p.I265V, c.793A>G; p.N189S, c.566A>G; and p.K103E, c.307A>G; none were present in dbSNP) insufficient to account for ARRP. Notably, an additional Japanese male with ARRP was found to carry the same frameshift variant p.L206fs, but heterozygously, with no other variants in the *NEK2* coding sequence. This same patient (R51) was later found to carry the retinitis pigmentosa GTPase regulator (*RPGR*) mutation c.2405_2406delAG; p.E802fs (Human Gene Mutation Database entry: CD004115), described previously to be a sufficient cause of RP (12).

In light of a recent study reporting the involvement of noncoding RNA in the pathogenesis of retinal degeneration in mice (13), variants in noncoding RNA were also analyzed. After the removal of variants observed in 52 publicly available control genomes, only isolated heterozygous variants each with one entry per gene remained, insufficient to account for ARRP.

***nek2* Inactivation and Rescue in Zebrafish.** To validate the pathogenic role of *NEK2* deficiency in RP, we suppressed the sole ortholog of *NEK2* in zebrafish embryos and asked whether this manipulation might give rise to photoreceptor phenotypes. Upon injection of 6 ng of *nek2* splice-blocking morpholino, we observed gross ocular defects, including microphthalmia and enlarged eye sockets in 5-d postfertilization (dpf) morphant (MO) embryos (Fig. 4*A*). Whereas 63% of MO embryos displayed such phenotypes, only 21% of embryos expressing both MO and wild-type human *NEK2* mRNA did, suggesting that the ocular phenotypes are specific to the *nek2* suppression ($P < 0.001$) (Fig. 4*B*).

We next asked whether, in addition to overt structural abnormalities that may not directly inform the involvement of this gene to RP in humans, suppression of *nek2* might also give rise to photoreceptor defects consistent with those of patients with ARRP. We therefore embedded and paraffin sectioned control and MO embryos. In addition to the small eye phenotype, we detected alterations in the photoreceptor layer. Specifically, after

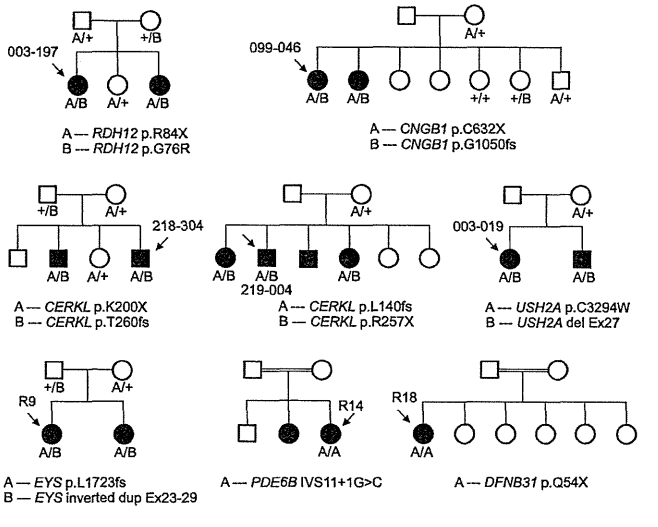

Fig. 3. Cosegregation analyses. All mutations analyzed cosegregated with the disease according to an autosomal recessive pattern of inheritance.

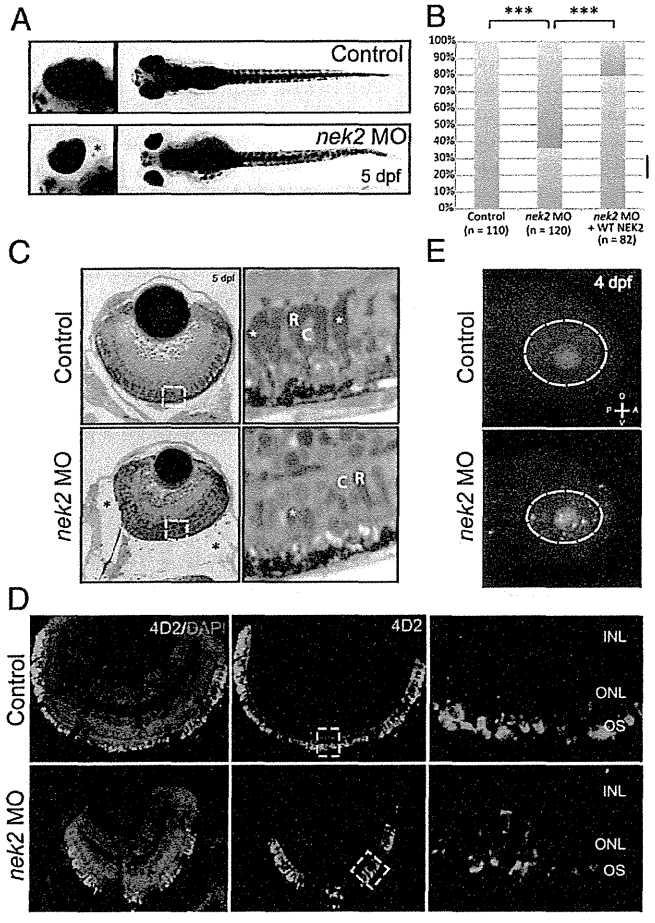

Fig. 4. In vivo functional evaluation of *nek2* loss in zebrafish. (*A*) Bright-field representation of 5-dpf control and *nek2* morphant zebrafish embryos. Magnified *Insets* highlight ocular phenotypes including microphthalmia and enlarged eye sockets (marked by the black asterisk). (*B*) Ocular phenotypes including microphthalmia and enlarged eye sockets vs. normal phenotypes (red bars and blue bars, respectively) are quantified in control and *nek2* morphant embryos, as well as in morphant animals rescued with human WT *NEK2* mRNA. Asterisks indicate statistically significant differences between groups ($P < 0.001$). (*C*) Histology of control and *nek2* morphant embryos also show enlarged eye sockets (marked by black asterisks) and microphthalmia. Magnified *Insets* show a decrease in the number of photoreceptors with apparent changes in domains of condensed chromatin (white asterisks). C, cones; R, rods. (*D*) Immunohistochemical analyses of retinal cryosections from control and *nek2* MO embryos, stained with DAPI (blue) and the 4D2 antibody against rhodopsin (green). Suppression of *nek2* results in the depletion of rods and in the mislocalization of rod opsin from the outer segment (OS) of photoreceptors. INL, inner nuclear layer; ONL, outer nuclear layer. (*E*) TUNEL immunofluorescent images of 4-dpf embryos, showing an increase in the number of apoptotic cells in *nek2* morphant embryos. The dotted ovals indicate the position of the eye. A, anterior; D, dorsal; P, posterior; V, ventral.

serial sectioning of 10–20 embryos injected with sham, MO, or MO + human *NEK2* mRNA, we observed a persistent decrease in the number of photoreceptors with large central domains of condensed chromatin. This phenotype was seen in all *nek2* MO embryos evaluated, but was absent from embryos injected with either sham or MO + human *NEK2* mRNA, suggesting a loss of rod photoreceptors specific to the suppression of *nek2* (Fig. 4*C* and *SI Appendix*, Fig. S5). To verify this observation, we used a rhodopsin (4D2) antibody to stain retinal cryosections from embryos injected with sham, MO, or MO + human *NEK2* mRNA (Fig. 4*D* and *SI Appendix*, Fig. S5). Immunohistochemical analyses of cross-sections from each condition demonstrated that the suppression of *nek2* resulted in the depletion of ~24% of 4D2-positive rod photoreceptors. In addition, mislocalization of rod opsin throughout the photoreceptor cells was evident in the central retina of *nek2* MO specimens, consistent with the hypothesis that nek2 is required for the appropriate trafficking of rhodopsin to the outer segments (Fig. 4*D*).

Further, to ask whether apoptosis, a major mechanism of photoreceptor loss in most known forms of RP (14), might account for some of the observed loss of photoreceptors, we performed TUNEL analysis. Masked scoring of embryos (~50 embryos per injection mixture) revealed a sevenfold increase in the number of TUNEL-positive cells in the eye and head region of *nek2* morphant embryos. By sharp contrast, we did not observe more than 1–10 TUNEL-positive cells in embryos injected with MO + *NEK2* mRNA (Fig. 4*E*).

Finally, we were intrigued by the discovery of a heterozygous frameshift variant p.L206fs in *NEK2* and the bona fide *RPGR* mutation p.E802fs in a patient with RP. We therefore asked whether the *RPGR* variant may interact genetically with the *NEK2* locus. To test this possibility, we coinjected subeffective doses of the *nek2* MO and *rpgr* MO and compared embryos with single or double MO ($n > 100$ at subeffective doses). Approximately 28% of embryos carrying subeffective doses of both *nek2* and *rpgr* MO revealed ocular ($P < 0.001$) and rod photoreceptor phenotypes (serial sectioning of 10 embryos per genotype) that exceeded the number of affected embryos induced by either *nek2* (3%) or *rpgr* (10%) MO alone, suggesting that the RPGR allele interacts *in trans* with the NEK2 locus to exacerbate photoreceptor defects (*SI Appendix*, Fig. S6).

Discussion

Massively parallel sequencing has proven to have a high potential to detect mutations in patients with rare Mendelian diseases (15). To date, most reports focus on monogenic conditions with no genetic heterogeneity, for which mutations can be recognized from benign variants since they invariantly affect the same gene in different patients.

In this study, we explored the efficacy of WGS in identifying mutations in unrelated patients from diverse ethnic backgrounds and presenting with a disease that is clinically the same but that has different genetic drivers. Whereas the small number of genomes analyzed in this study precludes an accurate analysis of quantitative measures, such as sensitivity of the WGS to detect mutations in known RP genes, we observed a few features that allowed us to make some valid comparisons between the different techniques currently available for genetic diagnosis. First, the majority of the pathogenic mutations identified were never reported before. This implies that tools that rely on systematic search for known pathogenic variants, both via mutation-centered resequencing and chip-based hybridization, may not be adequate for ARRP. Second, thanks to full-genome data, we detected complex structural variants whose junctions were located deep in noncoding regions. Because of their nature, these disease-causing variants would have been invisible to standard screening methods, or even to WES. Coverage-based analysis of CNV in exome sequencing has been attempted, with variable results. Limitations of this approach include the uneven efficiency of target DNA capture (and hence sequence coverage, on which assessment of number of copies is based) over different probes and, above all, the low probability of detecting junctions defining the SVs, which are more likely to be found in the nonexonic sequences composing ~98% of our genome. Unambiguous detection of abnormal junctions and mate pair information are crucial parameters in defining a SV; for instance, they allow distinguishing a tandem duplication from an inverted one. Third, because we had access to the full wealth of genomic information, we could integrate many sources of information

at once (e.g., SNP genotypes, phasing, etc.) that allowed us to accurately filter DNA variants that were related to the disease.

Genetic defects in *EYS* were proposed recently to be one of the major causes of ARRP in the Japanese population (16). We found that one of the pathogenic *EYS* alleles was a large SV (446 kb) with a complex genomic rearrangement. This finding supports the notion that SVs represent frequent pathogenic mutations in this gene (17). A homozygous nonsense mutation in exon 6 of *DFNB31* was identified in R18, a patient with nonsyndromic ARRP. The *DFNB31* gene encodes whirlin, a PDZ scaffold protein with expression in both hair cell stereocilia and retinal photoreceptor cells. Whirlin binds to the protein encoded by *USH2A* (18), a gene associated with both Usher syndrome type II (ARRP accompanied by hearing loss) and nonsyndromic ARRP (19). Whereas mutations in *DFNB31* have been reported as rare causes of Usher syndrome type II (20, 21), no DNA changes in its sequence have yet been associated with nonsyndromic ARRP. However, at the age of 66, the past medical history of this patient was significant for only hyperlipidemia and she did not report any hearing loss. We could not perform an auditory examination because she was no longer reachable.

In patients from consanguineous families, regions of IBD allowed restricting the search for pathogenic mutations to only a fraction of the genome. However, these same regions were susceptible to carrying other rare but nonpathogenic homozygous changes as well. Indeed, a higher number of candidate genes/mutations remained among Japanese patients with parental consanguinity compared with those without it (*SI Appendix*, Table S4). These results suggest that even if the analysis should be restricted to areas of IBD, genomes with high homozygosity do not necessarily offer an extra advantage in mutation detection, when comprehensive genomic sequencing in single individuals is performed.

In three patients we identified clear-cut pathogenic but heterozygous mutations in known ARRP genes that could not be associated directly with the disease. This was particularly evident for patient R14, who carried a heterozygous frameshift in *DFNB31* but was also homozygous for a mutation inactivating *PDE6B* (22). These findings are not surprising, given the elevated number of recessive ARRP mutations that are predicted to be present in the general population. Based both on theoretical assessments and on experimental data from control cohorts, we estimated that 1 in 3–7 individuals could be potential heterozygous carriers of an ARRP mutation (23, 24) or, as in the present case, 3 in 16.

The reasons why no candidate mutations of similar quality (i.e., two mutations, at least one of them being clearly deleterious in nature) to those revealed in known RP genes was uncovered in most of the unresolved genomes are unknown. Explanations for this observation may include the presence of variants or SVs that were undetected because of problems inherent in the mapping or sequencing procedure, or of less obvious pathogenic changes that alter splicing or transcription. These would include variants located in introns or in promoter regions, synonymous changes, or changes lying within important yet unannotated exons, genes, or genetic elements that have not been explored in the current study. Diseases caused by oligogenic modes of inheritance, or perhaps attributable to missense mutations for which efficient prioritization is difficult, is another possible explanation. De novo mutations in unknown dominant RP genes could also be evoked.

The search for mutations in unknown disease-causing genes revealed a number of genes with two nonsynonymous changes, which were mostly previously undescribed missenses. Application of more stringent filtering criteria by imposing the presence of at least one deleterious mutation followed by targeted annotation highlighted a single candidate, *NEK2*, in a Japanese patient who carried a homozygous frameshift in this gene. The serine/threonine-protein kinase NEK2 is known to play an important role in regulation of cell cycle progression through localization to the centrosomes and interaction with microtubules (25). The identified frameshift would result either in the creation of premature stop codon yielding a null allele or (less likely) a truncated protein lacking kinase activity and loss of microtubule binding. Importantly, defects in members of the Nek kinase family have been linked to impaired ciliogenesis and polycystic kidney disease (26). Recently, a role for *Nek2* in the left–right patterning of vital organs (a phenotype associated with ciliary function) was established in *Xenopus laevis* (27). In the same work, in situ hybridization revealed the expression of *nek2* transcripts in the eye (27). Furthermore, because NEK2 interacts with and can phosphorylate rootletin, a component of photoreceptor cilia (28, 29), *NEK2* was considered to be an important candidate for ARRP.

Our zebrafish studies showed that lack of Nek2 induces microphthalmia as a gross morphological phenotype. More importantly, in *nek2* morphants, we observed mistrafficking of rhodopsin, a hallmark of photoreceptor disease (30), and a reduced number of rod photoreceptors, likely via apoptotic processes. These phenotypes were rescued by injection of wild-type human *NEK2* mRNA, validating the specificity of the induced defects. Microphthalmia is a phenotype that is difficult to interpret in the present context but that is not uncommon to zebrafish models of RP (31, 32). Meanwhile, photoreceptor death, mistrafficking of rhodospsin, and reduction of the outer retinal layers are classical features of RP in both patients and animal models (7, 14, 33). Indeed, no microphthalmia was noted in patient R19.

Intriguingly, the *NEK2* frameshift identified in R19 was also present in R51, another patient with RP who had a deleterious mutation in *RPGR*. As the *RPGR* mutation in itself could explain the disease, an obvious question was whether the *NEK2* mutation might in fact represent a common benign allele. We therefore searched for this variant in 1,273 control Japanese individuals and found that none carried it (allele frequency $<3.9 \times 10^{-4}$). The p.L206fs mutation in *NEK2* is therefore exceedingly rare, such that its presence in a homozygous state in a patient is a strong argument in favor of its being an uncommon cause for ARRP. Although it is possible to attribute the presence of both *NEK2* and *RPGR* mutations in R51 to chance, a more parsimonious explanation is that mutations in these two genes, both expressed in the connecting cilium, act synergistically to define a severe RP phenotype, due to the established principles of mutational load and oligogenic interactions of pathogenic alleles (34). In turn, this would increase the likelihood for the patient of being examined at earlier ages and analyzed genetically. Multiple genetic modifier genes have been reported for cilia-encoding genes and especially for RPGR (35). These modifiers may account in part for the wide phenotypic spectrum associated with genetic defects in this gene, ranging from localized macular atrophy to retinitis pigmentosa of variable severity. To investigate the possibility of the cooperative effect between deficiencies in these two ciliary genes, we performed in vivo genetic interaction studies and showed that loss of Rpgr function can exacerbate Nek2 ocular phenotypes, including defects comprising the trapping of rhodopsin in the inner segment. Taken together, our genetic and functional data indicate that *NEK2* is a disease gene and that the retinal phenotype that results from its deficiency may represent a newly recognized ciliopathy.

To date, WGS has not been as widely explored as WES in the context of mutation detection. This can be attributed mainly to cost-related issues, because WGS is at least twice as expensive as WES procedures ensuring the same average coverage. We believe that the additional features displayed by WGS are worth the difference in price; however, this is a rather subjective matter that also depends on the disease that is being investigated. In the present case, WGS was essential to identify two pathogenic structural variations originating in introns. This is a significant finding, considering that only seven genomes could undergo SV analysis. Therefore, as a general rule, WGS is probably the strategy of

choice when detection of structural variants or mutations in noncoding regions represents an important element of investigation. In the long term, considering that costs associated with massively parallel sequencing technology is expected to fall further and that analysis pipelines continue to evolve, it is probable that WGS would be just as workable economically and physically as WES. Limitations of WGS include the requirement of high-quality DNA to explore the full leverage of the mate-pair mapping and the lack of reliable pipelines to detect SVs ranging in size from 50 to a few hundred base pairs. Unexpectedly, the difficulty accompanied by handling the large amount of data produced by WGS was not a significant obstacle, given the power of desktop computers presently available on the market. Whereas samples with suitable quality could be obtained through careful preparation of fresh DNA samples, under detection of SVs may be a more problematic issue to solve. This occurs because current mapping is based on two steps: mapping of the short reads aimed at detecting variations between 1 and 50 bp and mate-pair mapping for detection of SVs larger than a few hundred bases; to our knowledge, a solution that could fill the gap between these two mapping approaches remains to be found.

In conclusion, in this study we identified clear-cut causative mutations among the overwhelming number of DNA variants present in the human genome, in single patients from genetically diverse populations. This happened without ambiguities in a highly heterogeneous disease, ARRP, and in more than 50% of the individuals analyzed. Furthermore, two cases presented mutations involving noncoding parts of the genome. Considering that the majority of patients referred for molecular genetics diagnosis are isolated individuals, our results are relevant not only to basic research, but also to future clinical genetic testing.

Methods

Our research protocol involving humans and animals was approved by the institutional review boards of our respective universities and organizations. Written informed consent for providing medical information and blood samples was obtained from each patient. Experimental procedures are described in detail in *SI Appendix, Methods*.

ACKNOWLEDGMENTS. We thank Anna M. Siemiatkowska and Frans P. M. Cremers for sharing material from a person with RP, Adriana Ransijn for technical help, as well as Andrea Superti-Furga, and Luisa Bonafé for fruitful suggestions. Data storage was ensured by the Vital-IT Center for high-performance computing of the Swiss Institute of Bioinformatics. This work was supported by the Swiss National Science Foundation (Grant 310030_138346) and the Gebert Rüf Foundation, Switzerland (Rare Diseases-New Technologies Grant) (both to C.R.); a Center Grant from the Foundation Fighting Blindness (to E.L.B.); National Institutes of Health Grants DK072301 and MH-084018 (to N.K.); Ministry of Health, Labor and Welfare (MHLW) of Japan [Grant 23300101 (to S.I. and N. Matsumoto) and Grant 23300201 (to S.I.)]; MHLW, the Japan Science and Technology Agency, and the Strategic Research Program for Brain Sciences (N. Matsumoto); and a Grant-in-Aid for Scientific Research on Innovative Areas (transcription cycle) from the Ministry of Education, Culture, Sports, Science and Technology of Japan and the Takeda Science Foundation (to N. Matsumoto).

1. Maguire AM, et al. (2008) Safety and efficacy of gene transfer for Leber's congenital amaurosis. *N Engl J Med* 358(21):2240–2248.
2. Bainbridge JW, et al. (2008) Effect of gene therapy on visual function in Leber's congenital amaurosis. *N Engl J Med* 358(21):2231–2239.
3. Cideciyan AV, et al. (2008) Human gene therapy for RPE65 isomerase deficiency activates the retinoid cycle of vision but with slow rod kinetics. *Proc Natl Acad Sci USA* 105(39):15112–15117.
4. Berson EL (1993) Retinitis pigmentosa. The Friedenwald Lecture. *Invest Ophthalmol Vis Sci* 34(5):1659–1676.
5. Berson EL, Rosner B, Sandberg MA, Weigel-DiFranco C, Willett WC (2012) ω-3 intake and visual acuity in patients with retinitis pigmentosa receiving vitamin A. *Arch Ophthalmol* 130(6):707–711.
6. Berson EL, et al. (1993) A randomized trial of vitamin A and vitamin E supplementation for retinitis pigmentosa. *Arch Ophthalmol* 111(6):761–772.
7. Hartong DT, Berson EL, Dryja TP (2006) Retinitis pigmentosa. *Lancet* 368(9549): 1795–1809.
8. Jin ZB, et al. (2008) Identifying pathogenic genetic background of simplex or multiplex retinitis pigmentosa patients: A large scale mutation screening study. *J Med Genet* 45(7):465–472.
9. Tucker T, Marra M, Friedman JM (2009) Massively parallel sequencing: The next big thing in genetic medicine. *Am J Hum Genet* 85(2):142–154.
10. Tuson M, Marfany G, Gonzàlez-Duarte R (2004) Mutation of CERKL, a novel human ceramide kinase gene, causes autosomal recessive retinitis pigmentosa (RP26). *Am J Hum Genet* 74(1):128–138.
11. Aldahmesh MA, et al. (2009) Molecular characterization of retinitis pigmentosa in Saudi Arabia. *Mol Vis* 15:2464–2469.
12. Vervoort R, et al. (2000) Mutational hot spot within a new RPGR exon in X-linked retinitis pigmentosa. *Nat Genet* 25(4):462–466.
13. Sanuki R, et al. (2011) miR-124a is required for hippocampal axogenesis and retinal cone survival through Lhx2 suppression. *Nat Neurosci* 14(9):1125–1134.
14. Cottet S, Schorderet DF (2009) Mechanisms of apoptosis in retinitis pigmentosa. *Curr Mol Med* 9(3):375–383.
15. Rabbani B, Mahdieh N, Hosomichi K, Nakaoka H, Inoue I (2012) Next-generation sequencing: Impact of exome sequencing in characterizing Mendelian disorders. *J Hum Genet* 57(10):621–632.
16. Hosono K, et al. (2012) Two novel mutations in the EYS gene are possible major causes of autosomal recessive retinitis pigmentosa in the Japanese population. *PLoS ONE* 7(2):e31036.
17. Pieras JI, et al. (2011) Copy-number variations in EYS: A significant event in the appearance of arRP. *Invest Ophthalmol Vis Sci* 52(8):5625–5631.
18. van Wijk E, et al. (2006) The DFNB31 gene product whirlin connects to the Usher protein network in the cochlea and retina by direct association with USH2A and VLGR1. *Hum Mol Genet* 15(5):751–765.
19. Rivolta C, Sweklo EA, Berson EL, Dryja TP (2000) Missense mutation in the USH2A gene: Association with recessive retinitis pigmentosa without hearing loss. *Am J Hum Genet* 66(6):1975–1978.
20. Ebermann I, et al. (2007) A novel gene for Usher syndrome type 2: Mutations in the long isoform of whirlin are associated with retinitis pigmentosa and sensorineural hearing loss. *Hum Genet* 121(2):203–211.
21. Yang J, et al. (2010) Ablation of whirlin long isoform disrupts the USH2 protein complex and causes vision and hearing loss. *PLoS Genet* 6(5):e1000955.
22. McLaughlin ME, Sandberg MA, Berson EL, Dryja TP (1993) Recessive mutations in the gene encoding the beta-subunit of rod phosphodiesterase in patients with retinitis pigmentosa. *Nat Genet* 4(2):130–134.
23. Rivolta C, Sharon D, DeAngelis MM, Dryja TP (2002) Retinitis pigmentosa and allied diseases: Numerous diseases, genes, and inheritance patterns. *Hum Mol Genet* 11(10): 1219–1227.
24. Nishiguchi KM, Rivolta C (2012) Genes associated with retinitis pigmentosa and allied diseases are frequently mutated in the general population. *PLoS ONE* 7(7):e41902.
25. Fry AM, Meraldi P, Nigg EA (1998) A centrosomal function for the human Nek2 protein kinase, a member of the NIMA family of cell cycle regulators. *EMBO J* 17(2): 470–481.
26. Quarmby LM, Mahjoub MR (2005) Caught Nek-ing: Cilia and centrioles. *J Cell Sci* 118(Pt 22):5161–5169.
27. Fakhro KA, et al. (2011) Rare copy number variations in congenital heart disease patients identify unique genes in left-right patterning. *Proc Natl Acad Sci USA* 108(7): 2915–2920.
28. Bahe S, Stierhof YD, Wilkinson CJ, Leiss F, Nigg EA (2005) Rootletin forms centriole-associated filaments and functions in centrosome cohesion. *J Cell Biol* 171(1):27–33.
29. Yang J, et al. (2002) Rootletin, a novel coiled-coil protein, is a structural component of the ciliary rootlet. *J Cell Biol* 159(3):431–440.
30. Hollingsworth TJ, Gross AK (2012) Defective trafficking of rhodopsin and its role in retinal degenerations. *Int Rev Cell Mol Biol* 293:1–44.
31. Luo N, Lu J, Sun Y (2012) Evidence of a role of inositol polyphosphate 5-phosphatase INPP5E in cilia formation in zebrafish. *Vision Res* 75:98–107.
32. Patil SB, Hurd TW, Ghosh AK, Murga-Zamalloa CA, Khanna H (2011) Functional analysis of retinitis pigmentosa 2 (RP2) protein reveals variable pathogenic potential of disease-associated missense variants. *PLoS ONE* 6(6):e21379.
33. Chang GQ, Hao Y, Wong F (1993) Apoptosis: Final common pathway of photoreceptor death in rd, rds, and rhodopsin mutant mice. *Neuron* 11(4):595–605.
34. Davis EE, Katsanis N (2012) The ciliopathies: A transitional model into systems biology of human genetic disease. *Curr Opin Genet Dev* 22(3):290–303.
35. Fahim AT, et al. (2011) Allelic heterogeneity and genetic modifier loci contribute to clinical variation in males with X-linked retinitis pigmentosa due to RPGR mutations. *PLoS ONE* 6(8):e23021.

ARTICLE

De Novo Mutations in *GNAO1*, Encoding a $G\alpha_o$ Subunit of Heterotrimeric G Proteins, Cause Epileptic Encephalopathy

Kazuyuki Nakamura,[1,2,9] Hirofumi Kodera,[1,9] Tenpei Akita,[3,9] Masaaki Shiina,[4] Mitsuhiro Kato,[2] Hideki Hoshino,[5] Hiroshi Terashima,[5] Hitoshi Osaka,[6] Shinichi Nakamura,[7] Jun Tohyama,[8] Tatsuro Kumada,[3] Tomonori Furukawa,[3] Satomi Iwata,[3] Takashi Shiihara,[2,10] Masaya Kubota,[5] Satoko Miyatake,[1] Eriko Koshimizu,[1] Kiyomi Nishiyama,[1] Mitsuko Nakashima,[1] Yoshinori Tsurusaki,[1] Noriko Miyake,[1] Kiyoshi Hayasaka,[2] Kazuhiro Ogata,[4] Atsuo Fukuda,[3] Naomichi Matsumoto,[1,*] and Hirotomo Saitsu[1,*]

Heterotrimeric G proteins, composed of α, β, and γ subunits, can transduce a variety of signals from seven-transmembrane-type receptors to intracellular effectors. By whole-exome sequencing and subsequent mutation screening, we identified de novo heterozygous mutations in *GNAO1*, which encodes a $G\alpha_o$ subunit of heterotrimeric G proteins, in four individuals with epileptic encephalopathy. Two of the affected individuals also showed involuntary movements. Somatic mosaicism (approximately 35% to 50% of cells, distributed across multiple cell types, harbored the mutation) was shown in one individual. By mapping the mutation onto three-dimensional models of the Gα subunit in three different complexed states, we found that the three mutants (c.521A>G [p.Asp174Gly], c.836T>A [p.Ile279Asn], and c.572_592del [p.Thr191_Phe197del]) are predicted to destabilize the Gα subunit fold. A fourth mutant (c.607G>A), in which the Gly203 residue located within the highly conserved switch II region is substituted to Arg, is predicted to impair GTP binding and/or activation of downstream effectors, although the p.Gly203Arg substitution might not interfere with Gα binding to G-protein-coupled receptors. Transient-expression experiments suggested that localization to the plasma membrane was variably impaired in the three putatively destabilized mutants. Electrophysiological analysis showed that $G\alpha_o$-mediated inhibition of calcium currents by norepinephrine tended to be lower in three of the four $G\alpha_o$ mutants. These data suggest that aberrant $G\alpha_o$ signaling can cause multiple neurodevelopmental phenotypes, including epileptic encephalopathy and involuntary movements.

Introduction

Epileptic encephalopathy is a group of neurological disorders characterized by severe and progressive cognitive and behavioral impairments, which are most likely caused or made worse by epileptic activity.[1] Ohtahara syndrome (OS [MIM 308350 and 612164]) is the most severe and the earliest form of epileptic encephalopathy and is characterized by tonic spasms mainly in the neonatal period, seizure intractability, and a suppression-burst pattern on electroencephalography (EEG).[2] De novo mutations in three genes, *ARX* (MIM 300382), *STXBP1* (MIM 602926), and *KCNQ2* (MIM 602235), have been reported in individuals with OS.[3–6]

Heterotrimeric guanine-binding proteins (G proteins) are composed of α, β, and γ subunits. In its basal state, Gα is bound with guanosine diphosphate (GDP) and forms the Gαβγ complex. When a seven-transmembrane-type receptor binds its agonist, it activates G proteins by catalyzing the exchange of GDP for guanosine triphosphate (GTP) on the Gα subunit. Subsequently, GTP-bound Gα dissociates from Gβγ, and each of the two complexes activates distinct downstream effectors.[7] In mammals, Gα subunits are divided into four classes: $G\alpha_{i/o}$, $G\alpha_{q/11}$, $G\alpha_s$, and $G\alpha_{12/13}$.[7] $G\alpha_o$, encoded by *GNAO1* (MIM 139311), is extremely abundant in brain tissue, where it can constitute up to approximately 0.5% of membrane protein,[8] suggesting important roles in brain function. In fact, mice lacking $G\alpha_o$ show multiple neurological abnormalities, including generalized tremor, occasional seizures, severe motor-control impairment, hyperalgesia, and behavioral abnormalities with early postnatal lethality.[9,10]

In this study, de novo *GNAO1* mutations were identified in four epileptic-encephalopathy-affected individuals, three of whom were diagnosed with OS. In addition, two of the four individuals showed involuntary movements, suggesting that aberration of $G\alpha_o$ can cause multiple neurodevelopmental phenotypes.

[1]Department of Human Genetics, Yokohama City University Graduate School of Medicine, 3-9 Fukuura, Kanazawa-ku, Yokohama 236-0004, Japan; [2]Department of Pediatrics, Yamagata University Faculty of Medicine, 2-2-2 Iida-nishi, Yamagata 990-9585, Japan; [3]Department of Neurophysiology, Hamamatsu University School of Medicine, 1-20-1 Handayama, Higashi-ku, Hamamatsu 431-3192, Japan; [4]Department of Biochemistry, Yokohama City University Graduate School of Medicine, 3-9 Fukuura, Kanazawa-ku, Yokohama 236-0004, Japan; [5]Division of Neurology, National Center for Child Health and Development, 2-10-1 Okura, Setagaya-ku, Tokyo 157-8535, Japan; [6]Division of Neurology, Clinical Research Institute, Kanagawa Children's Medical Center, 2-138-4 Mutsukawa, Minami-ku, Yokohama 232-8555, Japan; [7]Department of Pediatrics, Nagano Red Cross Hospital, 5-22-1 Wakasato, Nagano 380-8582, Japan; [8]Department of Pediatrics, Epilepsy Center, Nishi-Niigata Chuo National Hospital, Niigata 950-2085, Japan
[9]These authors contributed equally to this work
[10]Present address: Department of Neurology, Gunma Children's Medical Center, 779 Shimohakoda Hokkitsu-machi, Shibukawa, Gunma 377-8577, Japan
*Correspondence: naomat@yokohama-cu.ac.jp (N.M.), hsaitsu@yokohama-cu.ac.jp (H.S.)
http://dx.doi.org/10.1016/j.ajhg.2013.07.014.

Subjects and Methods

Subjects

Twelve individuals with OS were previously analyzed by whole-exome sequencing (WES).[3,11] In addition, we analyzed parental samples from 5 of the 12 individuals by WES. Screening for *GNAO1* mutations was performed in 367 individuals with epileptic encephalopathy (including 62 OS cases) by high-resolution-melting (HRM) analysis (339 cases) and/or WES (100 cases). The diagnosis was made on the basis of clinical features and characteristic patterns on EEG. Experimental protocols were approved by the institutional review board of Yokohama City University School of Medicine and Yamagata University Faculty of Medicine. Informed consent was obtained from the families of all individuals.

DNA Samples

Genomic DNA was obtained from peripheral-blood leukocytes by standard methods. For detection of a mosaic mutation in individual 2, genomic DNA from saliva and nails was isolated with an Oragene DNA kit (DNA Genotek) and an ISOHAIR kit (Nippon Gene), respectively.

WES

Genomic DNA was captured with the SureSelect Human All Exon v.4 Kit (Agilent Technologies) and sequenced with four samples per lane on an Illumina HiSeq 2000 (Illumina) with 101 bp paired-end reads. Image analysis and base calling were performed by Sequencing Control Software with Real-Time Analysis and CASAVA software v.1.8 (Illumina). Exome data processing, variant calling, and variant annotation were performed as previously described.[12–14] Reads were aligned to GRCh37 with Novoalign (Novocraft Technologies). Duplicate reads were removed with Picard, and local realignments around indels and base-quality-score recalibration were performed with the Genome Analysis Toolkit (GATK).[13] Single-nucleotide variants and small indels were identified with the GATK UnifiedGenotyper and were filtered according to the Broad Institute's best-practice guidelines v.3. Not flagged as clinically associated, variants registered in dbSNP135 were filtered. Filter-passed variants were annotated with ANNOVAR.[14] Pathogenic mutations detected by WES were confirmed by Sanger sequencing.

Mutation Screening

Genomic DNA was amplified with an illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare). Exons 1–8 covering the *GNAO1* coding region of two transcript variants (transcript variant 1, RefSeq accession number NM_020988.2, encoding $G\alpha_{o1}$; transcript variant 2, RefSeq accession number NM_138736.2, encoding $G\alpha_{o2}$) were screened by HRM analysis. The last two exons differ between the transcript variants. HRM analysis was performed with a Light Cycler 480 (Roche Diagnostics). Samples showing an aberrant melting curve in the HRM analysis were sequenced. PCR primers and conditions are shown in Table S1, available online. All mutations not present in publically available databases were verified with original genomic DNA and were searched for in the variant database of our 408 in-house control exomes.

Deep Sequencing of a Mosaic Mutation

PCR products (length 178 bp) spanning the c.521A>G mutation were amplified with the use of blood, saliva, and nail DNA samples from individual 2 and blood DNA samples from her parents as a template. Adaptor ligation, nick repair, and amplification were performed with the Ion Xpress Plus Fragment Library Kit (Life Technologies) according to the manufacturer's protocol (part no. 4471989 Rev. B). Indexing was carried out with the Xpress Barcode Adapters 1–16 Kit (Life Technologies). Emulsion PCR and enrichment steps were carried out with the Ion OneTouch 200 Template Kit v.2 (Life Technologies) according to the manufacturer's protocol (part no. 4478371 Rev. A). Sequencing of the amplicon libraries was carried out on the Ion Torrent Personal Genome Machine (PGM) with the Ion 314 sequencing chip and the Ion PGM 200 Sequencing Kit (Life Technologies) according to the recommended protocol (part no. 4474246 Rev. B). Torrent Suite 2.2 was used for all analyses. The percentage of mosaicism was examined with the Integrative Genomics Viewer.[15,16]

Expression Vectors

A full-length human *GNAO1* cDNA clone (transcript variant 1, encoding $G\alpha_{o1}$) was purchased from Kazusa DNA Research Institute. Human *GNAO1* cDNA was inserted into a pEF6/V5-His-C vector for the introduction of a C-terminal V5 epitope (Life Technologies). Site-directed mutagenesis using a KOD-Plus-Mutagenesis kit (Toyobo) was performed for generating *GNAO1* mutants, including c.521A>G (p.Asp174Gly), c.572_592del (p.The191_Phe197del), c.836T>A (p.Ile279Asn), and c.607G>A (p.Gly203Arg). A c.607_609delinsACA (p.Gly203Thr) mutant, in which GTP binding was reversible in contrast to the WT,[17] was also generated to serve as the known loss-of-function mutant.[18] All variant cDNAs were confirmed by Sanger sequencing.

Immunofluorescence Microscopy

Mouse neuroblastoma 2A (N2A) cells were grown as previously described.[4] N2A cells on glass coverslips were transfected with 200 ng of plasmid DNA with the use of X-tremeGENE 9 DNA Transfection Reagent (Roche Diagnostics). After 24 hr, cells were fixed in PBS containing 4% paraformaldehyde for 15 min and permeabilized in PBS containing 0.1% Triton X-100 for 5 min. Cells were then blocked with 10% normal goat serum for 30 min. V5-tagged $G\alpha_{o1}$ was detected with a mouse V5 antibody (1:200 dilution; Life Technologies) and Alexa-488-conjugated goat anti-mouse IgG (1:1000 dilution; Life Technologies). Coverslips were mounted with Vectashield (Vector Laboratories) that contained DAPI and were visualized with an inverted FV1000-D confocal microscope (Olympus).

Structural Modeling and Free-Energy Calculations

We used FoldX software (version 3.0β5) to construct mutated molecular structures and calculate the free-energy changes caused by the mutations.[19] We used crystal structures of the GDP-bound inactive $G\alpha_i\beta\gamma$ heterotrimer (Protein Data Bank [PDB] 1GG2),[20] the nucleotide-free $G\alpha_s\beta\gamma$ in complex with agonist-occupied monomeric β2 adrenergic receptor (β2AR) (PDB 3SN6),[21] and the transition-state analog of GTP ($GDP^+AlF_4^-$)-bound $G\alpha_q$ in complex with its effector phospholipase C-β (PLCβ) (PDB 3OHM)[22] as three-dimensional structure models of the $G\alpha_o$ subunit in different complexed states. Each of the mutations, corresponding to p.Asp174Gly, p.Ile279Asn, or p.Gly203Arg in the human $G\alpha_o$ subunit, was introduced into the Gα subunit of each complex, and the free-energy change upon the mutation was calculated with FoldX software. Note that ligands included in the complexes were ignored in the calculation because the FoldX energy function

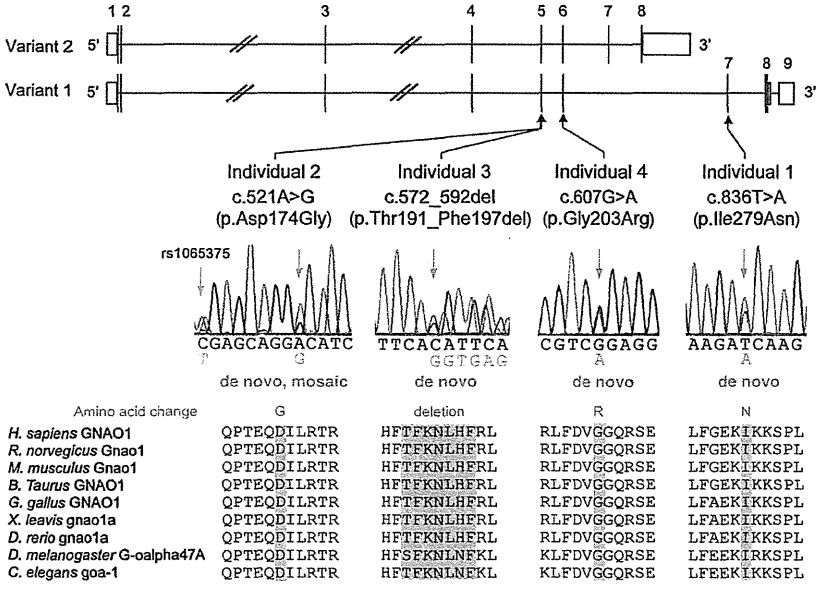

Figure 1. De Novo *GNAO1* Mutations in Individuals with Epileptic Encephalopathy
Schematic representation of *GNAO1*, including two transcript variants: transcript variant 1 (RefSeq NM_020988.2) with nine exons and transcript variant 2 (RefSeq NM_138736.2) with eight exons. The UTRs and coding regions are shown in white and black rectangles, respectively. Three mutations occurred in common exons of two transcript variants, and one mutation occurred uniquely in transcript variant 1. Note that the electropherogram of individual 2 suggested mosaicism of the c.521A>G mutation, and a heterozygous C>T change (rs1065375) was clearly demonstrated. All mutations caused substitution or deletion of evolutionarily conserved amino acids. Homologous sequences were aligned with the use of the CLUSTALW web site.

does not consider the contribution of ligands. The calculation was repeated three times, and the resultant data were presented as an average value with a SD.

Electrophysiology

For electrophysiological recording of calcium currents, we used NG108-15 cells transfected with individual *GNAO1* mutants. Expression vectors were introduced by electroporation with the Lonza Nucleofector device and the Cell Line Nucleofector Kit V (Lonza) according to the manufacture's protocol (program X-023). Two micrograms of plasmid DNA was used per transfection. The transfected cells were plated on poly-L-lysine-coated plastic coverslips (Cell Desk LF, MS-0113L; Sumitomo Bakelite) at a density of about 5×10^4 cells/cm^2 and cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS). One day after transfection, the cells were differentiated with DMEM supplemented with 10 μM prostaglandin E1, 50 μM IBMX, and 1% FBS for 3–7 days before recording. During the culture period, half of the medium was changed every other day.

The recording was made by the perforated whole-cell patch-clamp technique with amphotericin B. Cells on coverslips were perfused under an Olympus BX51W upright microscope (Olympus) with a bath solution containing 140 mM NaCl, 5 mM $CaCl_2$, 4 mM KCl, 1 mM $MgCl_2$, 10 mM HEPES, 10 mM TEACl, 8 mM glucose, and 0.0002 mM tetrodotoxin (pH 7.3 adjusted with NaOH). The patch pipette solution contained 100 mM CsCl, 10 mM EGTA, and 40 mM HEPES (pH 7.3 adjusted with CsOH). Amphotericin B was added to the pipette solution at 2 μl/ml just before the experiments. The pipettes were fabricated from borosilicate glass capillaries and had a resistance of 4–8 MΩ when backfilled with the amphotericin-B-containing pipette solution. The recording was started when the series resistance was reduced to <150 MΩ after gigaseal formation and clear cellular capacitive surges had appeared. Voltage-gated calcium currents were elicited by the application of 50 ms depolarizing pulses to +10 mV from the holding potential of −65 mV, recorded with a Multiclamp 700B (Molecular Devices) controlled via pCLAMP10 software (Molecular Devices), filtered at 2 kHz, and sampled at 10 kHz with 50% compensation for series resistance. $G\alpha_o$-mediated current inhibition was elicited by the application of 10 μM norepinephrine via the bath solution. After 3–5 min, inhibition was assessed by measurement of the changes in current density just before the end of the depolarizing pulses. Recordings were made at room temperature.

Statistical multiple comparisons were made with ANOVA followed by Dunnett's post hoc test, and the threshold p value for judging statistical significance was 0.05. The current inhibition induced by norepinephrine in individual mutant-expressing cells was assessed with a paired t test. The results are given as the mean ± SEM.

Results

GNAO1 Is Mutated in Individuals with Epileptic Encephalopathy

We previously performed WES of 12 individuals with OS.[3,11] In this study, we analyzed parental samples from 5 of the 12 individuals by WES (mean RefSeq read depth of 109) to systematically screen de novo or recessive mutations. We found no recessive mutations in *SLC25A22* (MIM 609302), *PNPO* (MIM 610090), *PNKP* (MIM 613402), *PLCB1* (MIM 613722), or *ST3GAL3* (MIM 615006), whose mutations were previously found in epileptic encephalopathy,[23–27] but we did find one or two de novo mutations in each of the five trio exomes. Among them, a de novo missense mutation (c.836T>A [p.Ile279Asn]) in *GNAO1* at 16q12.2 was identified in individual 1. In the exome data of the other seven original individuals, we also found in individual 2 a second missense mutation (c.521A>G [p.Asp174Gly]), which was confirmed as a de novo event by Sanger sequencing (Figure 1). Moreover, *GNAO1* mutation screening in 367 individuals with epileptic encephalopathy by HRM analysis (339 individuals) and/or WES (100 individuals,

Table 1. Clinical Features of Individuals with a *GNAO1* Mutation

	Individual 1	Individual 2	Individual 3	Individual 4
Gender	female	female	female	female
Age	13 years	4 years, 1 month	died at 11 months	8 years
Mutation	c.836T>A (p.Ile279Asn)	c.521A>G (p.Asp174Gly)	c.572_592 del (p.Thr191_Phe197 del)	c.607G>A (p.Gly203Arg)
Inheritance	de novo	de novo, somatic mosaic	de novo	de novo
Diagnosis	Ohtahara syndrome	Ohtahara syndrome	Ohtahara syndrome	epileptic encephalopathy
Initial symptom	tonic seizure at 4 days	series of tonic seizures at 29 days (tonic upgaze, head nodding, extension of all extremities)	series of tonic seizures at 2 weeks (resemble spasms)	opisthotonic posture, developmental delay at 7 months
Initial EEG	suppression-burst pattern at 4 days	suppression-burst pattern at 29 days	suppression-burst pattern at 2 weeks	diffuse irregular spike-and-slow-wave complex at 5 years
Course of seizures	tonic seizure at 5 years	series of tonic seizures at 9 months	tonic seizure at 10 months	focal seizure (tonic upgaze), tonic seizure at 5 years
Course of EEG	multifocal sharp waves at 1 year, 4 months; suppression-burst pattern at 5 years, 6 months	hypsarrhythmia at 3 months; diffuse spike-and-slow-wave complex at 1 year, 7 months; sharp waves at frontal lobe at 3 years, 9 months	hypsarrhythmia at 4 months	not done
Involuntary movement	-	-	dystonia	severe chorea, athetosis
Seizure control	intractable (2–3 times per day)	intractable (0–2 times per day)	intractable	intractable (several times per day)
Development				
Head control	-	+	-	-
Sitting	-	-	-	-
Meaningful words	-	-	-	-
MRI	normal at 1 month; cerebral atrophy at 5 years, 6 months	delayed myelination and thin corpus callosum at 10 months	normal at 3 months	delayed myelination at 1 year, 3 months; reduced cerebral white matter, thin corpus callosum at 4 years, 8 months

mean read depth of 129) revealed two de novo mutations: c.572_592 del (p.Thr191_Phe197del) in individual 3 and c.607G>A (p.Gly203Arg) in individual 4 (Figure 1). One mutation (c.836T>A) specifically affects *GNAO1* transcript variant 1, whereas the other three mutations affect both transcript variants 1 and 2. Web-based prediction tools suggested that these four mutations would be pathogenic (Table S2). None of the four mutations was found in the 6,500 exomes of the National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project Exome Variant Server or among our 408 in-house control exomes. Interestingly, exome data and Sanger sequencing indicated that the c.521A>G mutation in individual 2 was somatic mosaic (Figure 1 and Table S3). We confirmed de novo somatic mosaicism of the c.521A>G mutation by deep sequencing of PCR products amplified with blood, nail, and saliva DNA from individual 2 and blood DNA from her parents, showing that approximately 35%–50% of cells harbored the mutation (Table S3).

Phenotypes Associated with *GNAO1* Mutations

Neurological features of four female individuals with *GNAO1* mutations are shown in Table 1. Three individuals (individuals 1–3) developed tonic seizures with suppression-burst pattern on EEG at the onset (range 4–29 days), leading to a diagnosis of OS. Individuals 2 and 3 transited to West syndrome, a common infantile epileptic syndrome, as revealed by hypsarrhythmia on EEG at 3–4 months of age (Figures 2A–2C). Individual 4 showed developmental delay and opisthotonic posture at 7 months of age, and complex partial seizures with epileptic discharge on EEG was observed at 5 years (Figure 2D). Of note, two individuals showed involuntary movements: individual 3 showed dystonia, and individual 4 displayed chorea and athetosis (Table 1 and Movie S1). Brain MRI showed delayed myelination in individuals 2 and 4, cerebral atrophy or reduced cerebral white matter in individuals 1 and 4, and thin corpus callosum in individuals 2 and 4 (Figures 2E–2I). Although seizures and EEG

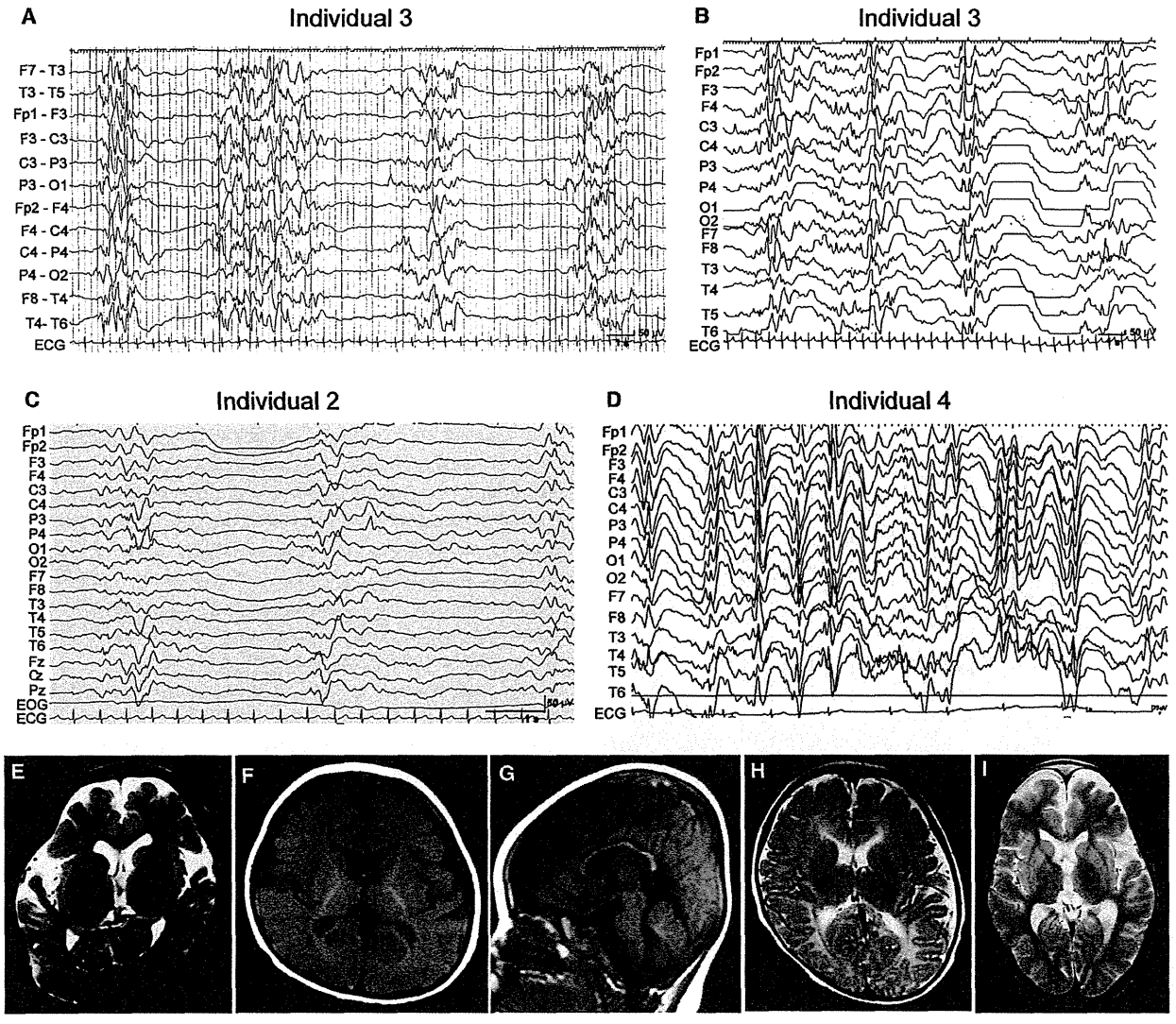

Figure 2. EEG and Brain MRI Features of Individuals with *GNAO1* Mutations
(A and B) Interictal EEG of individual 3. A suppression-burst pattern was observed at 2 months of age (A), and transition to hypsarrhythmia was observed at 4 months (B).
(C) Interictal EEG of individual 2 shows a suppression-burst pattern at 2 months.
(D) Interictal EEG of individual 4 shows a diffuse spike- or sharp-and-slow-wave complex at 5 years.
(E–I) T2-weighted axial images through the basal ganglia (E, H, and I) and T1-weighted axial (F) and sagittal (G) images. Individual 1 showed cerebral atrophy at 5 years and 6 months (E). Individual 2 showed delayed myelination and thin corpus callosum at 10 months (F and G). Individual 3 showed normal appearance at 3 months (H). Individual 4 showed reduced white matter at 7 years (I).

findings in two individuals with OS (individuals 2 and 3) were temporarily improved by adrenocorticotropic hormone therapy and valproic acid, all four individuals had intractable epileptic seizures in spite of combinatory therapy of antiepileptic drugs. All individuals had severe intellectual disability and motor developmental delay, and individual 3 died at 11 months because of respiratory-tract obstruction. These data suggest that *GNAO1* mutations can cause multiple neurodevelopmental phenotypes, including epileptic encephalopathy and involuntary movements.

Expression of Mutant $G\alpha_{o1}$ in N2A Cells

To examine the mutational effect of four *GNAO1* mutations, we performed transient expression experiments in N2A cells (Figure 3). C-terminally V5-epitope-tagged wild-type (WT) $G\alpha_{o1}$, encoded by transcript variant 1, was clearly localized in the cell periphery, as previously reported.[28] The p.Gly203Thr (with known loss of function)[17] and p.Gly203Arg (in individual 4) altered proteins were also localized in the cell periphery. In contrast, the p.Thr191_Phe197del altered protein (in individual 3) accumulated in the cytosolic compartment. The p.Asp174Gly

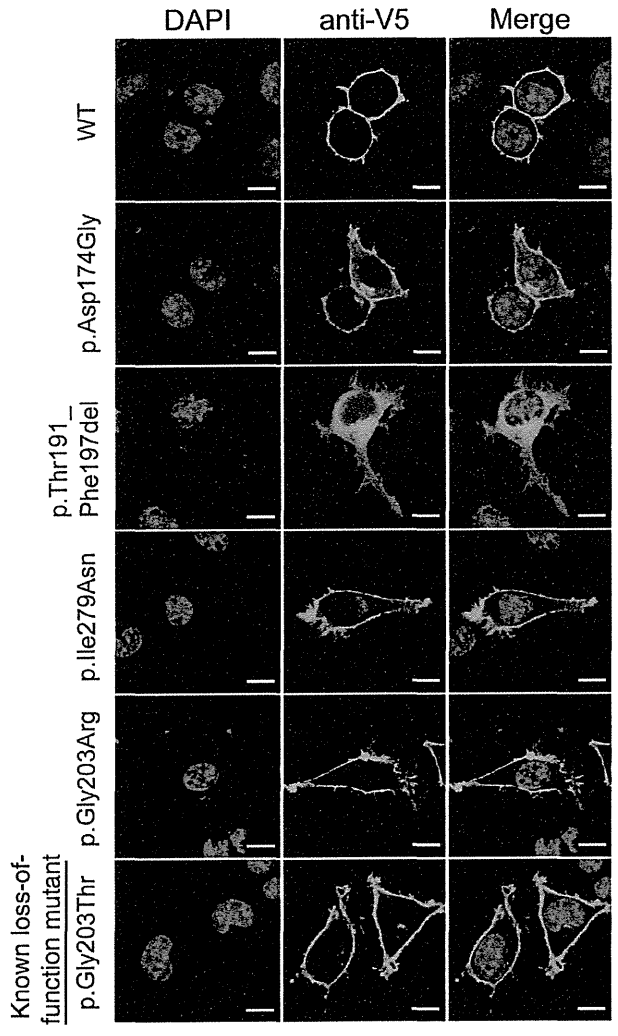

Figure 3. Localization of V5-Tagged $G\alpha_{o1}$ Proteins in N2A Cells
Localization of WT and five altered $G\alpha_{o1}$ proteins in N2A cells. The WT and p.Gly203Arg and p.Gly203Thr altered proteins were localized to the cell periphery. In contrast, the p.Thr191_Phe197del protein was localized to the cytosolic compartment. The other p.Asp174Gly and p.Ile279Asn proteins were localized to the cell periphery but were also observed in the cytosol. The scale bars represent 10 μm.

(individual 2) and p.Ile279Asn (individual 1) altered proteins were localized to the cell periphery and had weak signal in the cytosol, where more intense signal was observed in the p.Asp174Gly protein. Similar patterns of localization were observed for C-terminally AcGFP1-tagged $G\alpha_{o1}$ (Figure S1). These localization patterns suggest that the function of the p.Thr191_Phe197del altered protein might be most severely affected.

Structural Impacts of the Mutations on the Gα-Containing Complexes

To evaluate the impact of the *GNAO1* mutations on specific functions at the atomic level, we mapped the substituted positions onto structures of the Gα subunit in complexed states representing the GDP-bound inactive state, the nucleotide-free $G\alpha_s\beta\gamma$ in complex with the receptor, and the GTP-bound active state. In the case of point mutations, we further estimated free-energy changes of the mutations by using FoldX software (version 3.0β5).[19]

The region corresponding to amino acid residues 191–197 of human $G\alpha_{o1}$ is located in β strands and their connecting loop region and is involved in interactions with the G-protein-coupled receptor (GPCR) in the $G\alpha\beta\gamma$-β2AR complex (Figure 4A and Figure S2A). Thus, the deletion would affect secondary structure of the molecule and would not only impair the interaction with GPCR but also severely destabilize the Gα-subunit fold. The substituted residues corresponding to Asp174 and Ile279 of the human $G\alpha_{o1}$ subunit are both buried inside the protein (Figure 4A) and are involved in hydrogen-bonding and hydrophobic interactions, respectively (Figure S2B). Therefore, the p.Asp174Gly and p.Ile279Asn substitutions would destabilize the Gα-subunit fold, as supported by FoldX calculations showing a more than 2 kcal/mol increase in free-energy changes for these substitutions (Figure 4B). It can be speculated that these altered proteins tend to be misfolded or denatured in N2A cells and thus have altered cellular localization (Figure 3).

The substituted residue corresponding to Gly203 of human $G\alpha_{o1}$ is located within the highly conserved switch II region, responsible for activation of downstream effectors upon GTP binding (Figure 4A). Conformations of the switch regions differ depending on the complex state of the G protein. In the $G\alpha\beta\gamma$ heterotrimer and the GDP^+ AlF_4^--bound Gα-effector (PLCβ) complex, the glycine residues are closely surrounded by the switch I region and GTP (Figure 4A and Figure S2C). Thus, the p.Gly203Arg substitution would cause steric hindrance between the arginine side chain and the switch I region and/or GTP, destabilizing the complex, as supported by the FoldX calculations showing a remarkable increase in free-energy change upon the p.Gly203Arg substitution. By contrast, in the $G\alpha\beta\gamma$-receptor (β2AR) complex, no substantial steric hindrance was predicted from the structural modeling and FoldX calculations (Figure 4B and Figure S2C). These findings suggest that the p.Gly203Arg-substituted Gα subunit would impair GTP binding and/or activation of the downstream effectors, although it might still bind to GPCR. This prediction was supported by previous reports, in which GTP binding was weakened in the p.Gly203Thr altered Gα.[17] This also appears to be consistent with the apparently normal cellular localization of the p.Gly203Arg altered protein in N2A cells (Figure 3).

Electrophysiological Evaluation of $G\alpha_{o1}$ Mutants

It has been reported that N-type calcium channels are inhibited, at least in part, via $G\alpha_o$-mediated signaling.[7] Using NG108-15 cells, in which norepinephrine-induced calcium-current inhibition is mediated by $G\alpha_o$ (Figure 5A),[29] we analyzed functional properties of altered $G\alpha_{o1}$. Compared with cells expressing WT $G\alpha_{o1}$ (the

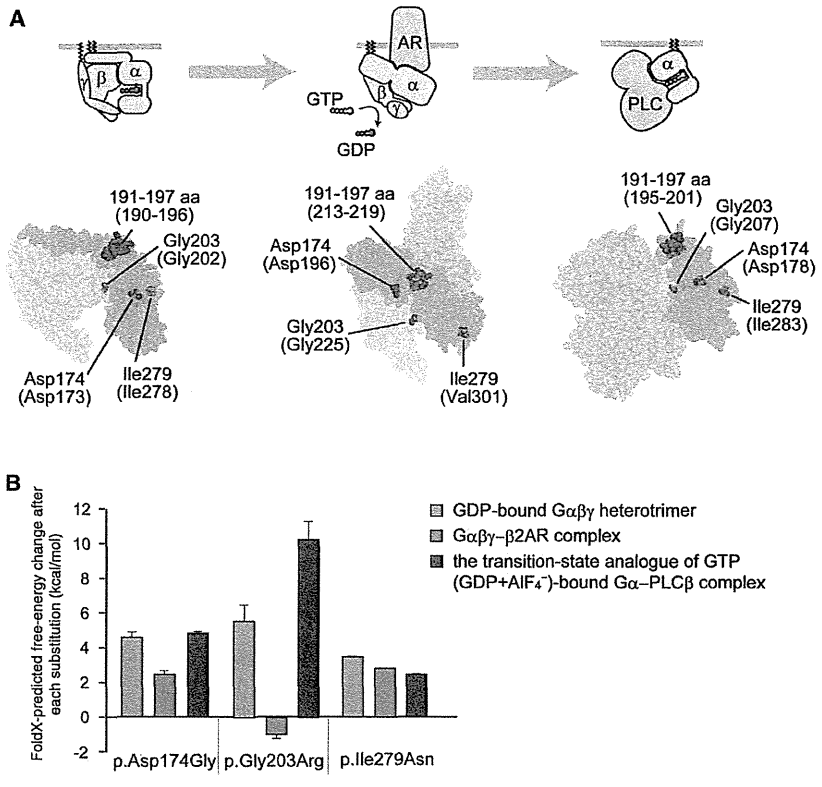

Figure 4. Structural Consideration of the Gα Amino Acid Substitutions in Some Complexed States

(A) Map of the amino acid substitution sites on the crystal structures of Gα-containing complexes: the GDP-bound inactive $G\alpha_i\beta\gamma$ heterotrimer (left), the nucleotide-free $G\alpha_s\beta\gamma$ in complex with an agonist-occupied monomeric β2AR (center), and the $GDP^+AlF_4^-$-bound $G\alpha_q$ in complex with its effector PLCβ (right). Molecular structures are shown as space-filling representations (from PyMOL). Gα, Gβ, and Gγ subunits are colored green, yellow, and pink, respectively, and the switch I and switch II regions in the Gα subunit are in light green. The β2AR (center) and PLCβ (right) molecules are colored brown. The substituted sites are shown in red, and the indicated amino acid numbers correspond to human $G\alpha o_1$ and, in parentheses, rat $G\alpha i_1$ (UniProtKB/Swiss-Prot P10824) (left), bovine Gαs (UniProtKB/Swiss-Prot P04896) (center), and mouse Gαq (UniProtKB/Swiss-Prot P21279) (right). The illustrations above each model show the orientation of each subunit and the bound molecules.

(B) The free-energy change after each of the amino acid substitutions was estimated from calculations using FoldX software. Each error bar represents an average value with a SD.

leftmost column in Figure 5B), NG108-15 cells expressing the p.Thr191_Phe197del substitution revealed a significant increase in calcium-current density before application of norepinephrine ($p < 0.05$ by Dunnett's post hoc test; the second column from the right in Figure 5B), suggesting that localization of the altered $G\alpha_{o1}$ might affect calcium-channel activity. In cells expressing the p.Asp174Gly substitution, a mild increase in the current density was also suggested, although the difference was not significant (the third column from the left in Figure 5B). The other two substitutions had no effects on the current (the second column from the left and the rightmost column in Figure 5B). Treatment with 10 μM norepinephrine reduced the calcium-current density by 19.0% ± 5.0% in cells expressing WT $G\alpha_{o1}$ ($p < 0.01$ by paired t test; left panel in Figure 5A and the leftmost bar in Figure 5C). A similar reduction was observed in cells expressing the p.Ile279Asn alteration (18.5% ± 3.5%, $p < 0.01$ by paired t test; the rightmost bar in Figure 5C). In cells expressing the p.Thr191_Phe197del alteration, by contrast, the reduction was obscured (12.1% ± 5.0%, not significant by paired t test; right panel in Figure 5A and the second bar from the right in Figure 5C). In cells expressing the other two substitutions (p.Gly203Thr and p.Asp174Gly), weaker current inhibition by norepinephrine was suggested (9.9% ± 3.8% and 11.1% ± 3.5%, respectively; both were $p < 0.05$ by paired t test; the second and third bars from the left in Figure 5C), although compared with that in WT-expressing cells, the degrees of inhibition in Gly203Thr- and p.Asp174Gly-expressing cells did not reach statistical significance (not significant by ANOVA). These data suggest that *GNAO1* mutations could hamper $G\alpha_o$-mediated signaling.

Discussion

We successfully identified four de novo heterozygous missense *GNAO1* mutations in four individuals. All four individuals showed severe intellectual disability and motor developmental delay, demonstrating that aberration of $G\alpha_o$ affects intellectual and motor development. In addition, all four individuals showed epileptic encephalopathy, and two of them showed involuntary movements. Because $G\alpha_o$-deficient mice show occasional seizures and generalized tremor,[9,10] it is likely that epilepsy and involuntary movement are two of the characteristic features caused by *GNAO1* mutations. Although $G\alpha_o$-deficient mice also show hyperactivity and hyperalgesia,[10] it is difficult to evaluate whether our individuals had these symptoms because of severe motor and cognitive impairment.

All four of these mutations, and especially two mutations leading to the p.Thr191_Phe197del and p.Gly203Arg substitutions, are predicted to affect $G\alpha_o$ function by structural evaluation. In fact, transient expression in N2A cells showed that localization of the p.Thr191_Phe197del altered protein was dramatically changed to the cytosolic compartment. Interestingly, two alterations (p.Ile279Asn