Journal of Human Genetics (2012) 57, 621–632

www.nature.com/jhg

REVIEW

Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders

Bahareh Rabbani[1], Nejat Mahdieh[1], Kazuyoshi Hosomichi, Hirofumi Nakaoka and Ituro Inoue

Traditional approaches for gene mapping from candidate gene studies to positional cloning strategies have been applied for Mendelian disorders. Since 2005, next-generation sequencing (NGS) technologies are improving as rapid, high-throughput and cost-effective approaches to fulfill medical sciences and research demands. Using NGS, the underlying causative genes are directly distinguished via a systematic filtering, in which the identified gene variants are checked for novelty and functionality. During the past 2 years, the role of more than 100 genes has been distinguished in rare Mendelian disorders by means of whole-exome sequencing (WES). Combination of WES with traditional approaches, consistent with linkage analysis, has had the greatest impact on those disorders following autosomal mode of inheritance; in more than 60 identified genes, the causal variants have been transmitted at homozygous or compound heterozygous state. Recent literatures focusing on identified new causal genes in Mendelian disorders using WES are reviewed in the present survey.
Journal of Human Genetics (2012) **57,** 621–632; doi:10.1038/jhg.2012.91; published online 26 July 2012

Keywords: exome sequencing; mendelian disorder; mutation; next-generation sequencing; NGS; WES

INTRODUCTION

One of the major concerns of medical sciences is finding the causal genes underlying human diseases. New technologies are developed, and progress in elucidating genetic basis of disorders is now one of the most discussing topics in medical genomics. With the advent of next-generation sequencing (NGS) technology, identification of genetic variations that serve as disease causality is progressing at rapid pace, which would improve the disease management either by available treatments or genetic counseling for the children's health and risk assessment of the relatives.[1] Molecular diagnosis, carrier detection, prenatal diagnosis and developing new therapies are the concern for the health care and society. Besides, unraveling the susceptible variants potentiates new interventions and prevention of the disease caused by different risk factors.

Generally, genetic disorders are categorized into monogenic and multifactorial disorders. Monogenic (single gene) disorders include simple and rare disorders. Multifactorial disorders comprise of complex disorders, multiple genes, as well as lifestyle or environmental factors, that are contributed to the disease. Rare genetic disorders have low prevalence of about at most 6.5 out of every 10 000 individuals according to the World Health Organization.[2–5] Tremendous efforts are now being performed for understanding the rare monogenic and complex traits and manifesting the genetic basis of the disease based on exome (all exons in the genome) sequencing.

Up to 27 January 2012, a total of 21 058 entries were reported in OMIM (Online Mendelian Inheritance in Man),[6] describing 13 790 genes and 4535 disorders with known molecular basis (http://omim.org/statistics/geneMap). Approximately, 1800 entries had phenotypic descriptions or known loci with unknown molecular bases and 2000 entries have been stated based on suspected Mendelian basis, and mainly the phenotypes are known. Bearing in mind, the Mendelian disorders that explore the novel genetic mechanisms, phenotypic variability, modifier genes, allelic variations and genetic variations of the diseases, may also provide clues in understanding the complex disorders.[7,8] Here, we focus on those diseases that are caused by single genes which their causal variants have been explored by exome sequencing. Sporadic cases are also included in this survey.

FROM DNA STRUCTURE REPORT TO NEW NGS REPORTS

Nearly 10 years after the discovery of DNA structure, the first gene was completely sequenced.[9] In 1977, Sanger *et al.*[10] and Maxam and Gilbert[11] developed initial sequencing methods (Table 1); meanwhile, the majority of human DNA sequence data has been described using the Sanger sequencing and fluorescence-based electrophoresis technologies. With the development of a revolutionary method named PCR by Kary Mullis in the 1980s, molecular genomic field has undergone enormous advances. As a matter of fact, a growing variety of molecular methods, including high-throughput sequencing technologies, has been emerged over the past 7 years. In 2004, the second (next)-generation sequencing methods, massively parallel sequencing platforms, were introduced and the next revolution in molecular genetics, including finding the disease-causing genes, is expected.

Division of Human Genetics, National Institute of Genetics, Mishima, Japan
[1]These authors contributed equally to this work.
Correspondence: Professor I Inoue, Division of Human Genetics, National Institute of Genetics, 1111 Yata, Shizuoka, Mishima 411-8540, Japan.
E-mail: itinoue@nig.ac.jp
Received 6 April 2012; revised 29 June 2012; accepted 29 June 2012; published online 26 July 2012

Table 1 Landmark events from DNA structure identification to new NGS reports

Year	*Event*	*Reference*
1953	Watson and Crick infer DNA's structure	Watson and Crick[65]
1964	The first nucleotide sequence of the gene encoding yeast alanine tRNA was reported	Holley *et al.*[9]
1977	Initial DNA sequencing methods were introduced by Sanger, Maxam and Gilbert	Sanger *et al.*[10] Maxam and Gilbert[11]
1980	First human linkage map based on restriction fragment length polymorphism	Botstein *et al.*[66]
1983	First dominant disease locus on the basis of linkage	Gusella *et al.*[15]
1985	Mullis discovered PCR technique	Mullis *et al.*[67]
1986	The idea of human genome sequencing was proposed	Smith *et al.*[68]
	The first human disease gene was cloned	Royer–Pokora*et al.*[69]
1987	The first homozygosity mapping was done	Lander and Botstein[16]
1989	First positional cloning of a recessive disease gene on the basis of linkage	Riordan *et al.*[14]
1993	A first-generation physical map of the human genome	Cohen *et al.*[70]
1995	First-genome sequence of an organism (*Hemophilus influenza)* was reported	Fleischmann *et al.*[71]
1999	First human chromosome was sequenced	Dunham *et al.*[72]
2000	Fruit fly genome was sequenced	Adams *et al.*[73]
	First assembly of the human genome was completed	Myers *et al.*[74]
2001	The first draft of human genome sequence was published	Venter *et al.*[75] Lander *et al.*[76]
2003	The human genome sequence was completed	Jasny and Roberts 2003
2004	Massively parallel sequencing platforms giving rise to the 'next-generation sequencing' were introduced	http://www.genome.gov/12513210
2005	The first NGS instrument was on market	Margulies *et al.*[77]
2008	First individual genome based on NGS was published	Wheeler *et al.*[78]
2009	Proof of principle: disease-gene identification by WES	Ng *et al.*[79]
2010	The first successful application of WES to identify the gene for a rare Mendelian disorder	Ng *et al.*[12]

Most of these platforms rely on sequencing by synthesis, and generation of their clonally clustered amplicons are achieved mainly through *in situ* polonies, emulsion PCR or bridge PCR. NGS technology affords high speed and throughput, both qualitative and quantitative sequence data, equivalent to the data from human genome project, in 10–20 days. Several different ways are employed in which NGS is being applied for identifying causal gene variant in the rare diseases. Whole-genome sequencing (WGS), whole-exome sequencing (WES), transcriptome sequencing, methylome and other sequencing approaches are applied in NGS systems.

Miller syndrome is the first rare Mendelian disorder from which its causal variants were identified, owing to the development of WES.[12] These researchers explained *DHODH* mutations in three affected pedigrees after filtering against public single-nucleotide polymorphism (SNP) databases and eight HapMap exomes.

There are increasing numbers of reports identifying the causal variants of the diseases. More than 100 causative genes in various Mendelian disorders have been identified by means of exome sequencing. Up to May 2012, a PUBMED search on Mendelian disorders using exome sequencing revealed 102 diseases as summarized in Table 2. A total of 326 exomes have been reported in these successful applications of WES in identifying novel causative genes. In all, 61 out of these 108 identified genes (56.5%) follow autosomal recessive, 40 of them are transmitted as autosomal dominant (37%), one case is X-linked recessive and one case follows X-linked dominant inheritance. About 35.2% of these genes (38 out of 108) have been identified by WES with only one exome, which was confirmed by sequencing of the possible identified genes in other patients. In overall, about three exomes (108/326) might be enough to be sequenced for identification of causative gene of the disease (Table 2). Evidently, we have no information about the unsuccessful WES studies. In addition to gene discovery of dominant and recessive diseases, WES has been used for determining somatic mutations in tumors and rare mutations with moderate effect in common diseases as well as clinical diagnoses.[13] As mentioned previously, here, we focus on the single-gene disorders.

EXOME SEQUENCING AND IDENTIFYING CAUSAL VARIANTS

Traditionally, the single-gene disorders were first analyzed based on linkage analyses[14,15] followed by positional cloning; an informative segregation pattern, clear mode of inheritance and enough affected family members could support for gene identification. Homozygosity mapping establishes loci of autosomal recessive disorders.[16] More complex forms of single-gene disorders, such as retinitis pigmentosa[17] and hearing loss, with different inheritance modes have been reported based on SNP arrays. Allelic association studies of case-control design are suitable for identifying highly associated SNPs with the complex diseases.

Drawbacks and limitations to these approaches that hindered the gene discovery need to be emphasized; there are families with small number of affected individuals, which do not meet the criteria required for classical gene-discovery methods. In addition, finding the causal genes in families fitting the criteria is very difficult in case of expression variability, locus heterogeneity, phenotypic heterogeneity, reduced penetrance or reduced fitness, because in these conditions, the causal effect could hardly be co-segregated with affection status within the family. Exome sequencing permits to overcome these obstacles. Also, there may be several sporadic cases from different families with similar phenotypes, in which exome sequencing interrogates the causal variants. As in Table 2, WES could identify the causal variants with a limited number of patients. Indeed, NGS technologies bring us new sights in unraveling the genetic basis of diseases.

Table 2 Mendelian disease-gene identifications by exome sequencing

Year	Disorder	MI	Location	Gene	Number exome sequenced	Sequencer	Capturing method	References
2010								
1	Miller syndrome	AR	16q22	*DHODH*	4	GAII[a]	Array	Ng *et al.*[12]
2	Autoimmune lymphoproliferative syndrome	AR	11q13.3	*FADD*	1	ND	SS[b]	Bolze *et al.*[80]
3	Nonsyndromic hearing loss	AR	1p13.3	*GPSM2*	1	GAIIx	SS	Walsh *et al.*[81]
4	Combined hypolipidemia	AR	1p31.1–p22.3	*ANGPTL3*	2	GAII	S	Musunuru *et al.*[37]
5	Perrault syndrome	AR	5q21	*HSD17B4*	1	GAIIx	SS	Pierce *et al.*[38]
6	Complex I deficiency	AR	3q21.3	*ACAD9*	1	SOLiD[c]	SS	Haack *et al.*[82]
7	Hyperphosphatasia mental retardation syndrome	AR	1p36.11	*PIGV*	3	SOLiD	SS	Krawitz *et al.*[83]
8	Sensenbrenner syndrome	AR	2p24.1	*WDR35*	2	SOLiD	SS	Gilissen *et al.*[55]
9	Cerebral cortical malformations	AR	19q13.12	*WDR62*	3	GAIIx	N[d] solid array	Bilguvar *et al.*[84]
10	3MC syndrome	AR	3q27-q28	*MASP1*	2	GAIIx	SS	Sirmaci *et al.*[47]
11	Kabuki syndrome	AD	12q13.12	*MLL2*	10	GAII	Array	Ng *et al.*[39]
12	Schinzel–Giedion syndrome	AD	18q21.1	*SETBP1*	4	SOLiD	SS	Hoischen *et al.*[53]
13	Spinocerebellar ataxia	AD	20p13	*TGM6*	4	GAII	NA	Wang *et al.*[44]
14	Terminal osseous dysplasia	XLD	Xq28	*FLNA*	2	GAII	SS	Sun *et al.*[45]
2011								
15	Nonsyndromic mental retardation	AR	19p13.12	*TECR*	6	GAII	S	Caliskan *et al.*[85]
16	Retinitis pigmentosa	AR	1p36.11	*DHDDS*	4	GAII	N	Züchner *et al.*[86]
17	Osteogenesisimperfecta	AR	17p13.3	*SERPINF1*	1	SOLiD 4	SS	Becker *et al.*[87]
18	Skeletal dysplasia	AR	8q22.1	*POP1*	4	GAII	NS	Glazov *et al.*[88]
19	Combined malonic and methylmalonicaciduria	AR	16q24.3	*ACSF3*	1	GAIIx	SS	Sloan *et al.*[89]
20	Amelogenesis	AR	17q24.2	*FAM20A*	1	SOLiD	SS	O'Sullivan *et al.*[90]
21	Chondrodysplasia and abnormal joint development	AR	8q12.1	*IMPAD1*	3	SOLiD 4	SS	Vissers *et al.*[91]
22	Progeroid syndrome	AR	11q13.1	*BANF1*	3	GAIIx	SS	Puente *et al.*[43]
23	Infantile mitochondrial cardiomyopathy	AR	6p21.1	*AARS2*	1	GAIIx	SS	Götz *et al.*[92]
24	Heterotaxy	AR	4q21.1	*SHROOM3*	1	GAII	NS	Tariq *et al.*[93]
25	Mosaic variegated aneuploidy syndrome	AR	11q21	*CEP57*	2	GAIIx	SS	Snape *et al.*[94]
26	Hypomyelinatingleukoencephalopathy	AR	10q22-q23 and 12q23.3	*POLR3A and POLR3B*	3	GAIIx	SS	Saitsu *et al.*[54]
27	Spastic ataxia-neuropathy syndrome	AR	18p11	*AFG3L2*	4	GAIIx	SS	Pierson *et al.*[42]
28	Dilated cardiomyopathy	AR	7q21-q22	*GATAD1*	2	GAIIx	SS	Theis *et al.*[95]
29	Gonadal dysgenesis	AR	17q21.2	*PSMC3IP*	2	SOLiD 3	SS	Zangen *et al.*[41]
30	Autosomal recessive progressive external ophthalmoplegia	AR	8q23.1	*RRM2B*	1	GAII	SS	Takata *et al.*[96]
31	Knobloch syndrome	AR	16q23	*ADAMTS18*	1	HiSeq2000	SS	Aldahmesh *et al.*[97]
32	Spinocerebellar ataxia with psychomotor retardation	AR	1q32.2	*SYT14*	2	GAIIx	SS	Doi *et al.*[98]
33	Adams–Oliver syndrome	AR	19p13.2	*DOCK6*	4	HiSeq2000	T	Shaheen *et al.*[99]
34	Steroid-resistant nephrotic syndrome	AR	15q21-q22 and 15q24.2	*MYO1E and NEIL1*	1	SOLiD4	SA	Sanna-Cherchi *et al.*[100]
35	Complex bilateral occipital cortical gyration abnormalities	AR	9q31-q34	LAMC3	1	GAIIx	NA	Barak *et al.*[101]
36	Intellectual disability	AR	14q12, 1p13.2 15q21.2	*AP4S1, AP4B1, AP4E1*	1	GAIIx	SS	Abou Jamra *et al.*[50]
37	Hypertrophic cardiomyopathy	AR	3q21-q23	*MRPL3*	1	GAIIx	SS	Galmiche *et al.*[102]
38	Retinitis pigmentosa	AR	6p24	*MAK*	1	GAIIx	SS	Ozgül *et al.*[103]
39	3M syndrome	AR	19q13.32	*CCDC8*	3	SOLiD 4	SS	Hanson *et al.*[104]
40	Seckel syndrome	AR	15q21.1	*CEP152*	1	GAII	SS	Kalay *et al.*[105]
41	ADK deficiency	AR	10q11-q24	*ADK*	2	HiSeq2000	SS	Bjursell *et al.*[106]
42	Nephronophthisis-like nephropathy	AR	4p14	*WDR19*	1	SOLiD	SS	Bredrup *et al.*[107]
43	Psuedo–Sjögren–Larsson syndrome	AR	6q14	*ELOVL4*	1	HiSeq2000	ND	Aldahmesh *et al.*[108]
44	Idiopathic infantile hypercalcemia	AR	20q13	*CYP24A1*	1	GAII	S	Dauber *et al.*[109]

Table 2 (Continued)

Year	Disorder	MI	Location	Gene	Number exome sequenced	Sequencer	Capturing method	References
45	EMARDD	AR	5q33	*MEGF10*	1	GAIIx	SS	Logan *et al.*[110]
46	Gray platelet syndrome	AR	3p21.31	*NBEAL2*	4	GAII	SS	Albers *et al.*[111]
47	Immunodeficiency, centromeric instability and facial anomalies	AR	6q21	*ZBTB24*	1	GAIIx	NA	deGreef *et al.*[112]
48	Leber congenital amaurosis	AR	2q37	*KCNJ13*	1	HiSeq2000	SS	Sergouniotis *et al.*[113]
49	Hereditary spastic paraparesis	AR	2q37.3	*KIF1A*	5	illumina	array	Erlich *et al.*[114]
50	Ohdo syndrome	AD	10q22.2	*KAT6B*	4	SOLiD 4	SS	Clayton-Smith *et al.*[115]
51	Paroxysmal kinesigenicdyskinesias	AD	16p11.2	*PRRT2*	4	HiSeq2000	SS	Chen *et al.*[116]
52	Hajdu–Cheney syndrome	AD	1p13-p11	*NOTCH2*	3	GAIIx	SS	Simpson *et al.*[117]
					6	GAIIx	SA	Isidor *et al.*[118]
53	Bohring–Opitz syndrome	AD	20q11	ASXL1	3	SOLiD4	SS	Hoischen *et al.*[119]
54	Hereditary diffuse leukoencephalopathy with spheroids	AD	5q32	*CSF1R*	2	HiSeq2000	SS	Rademakers *et al.*[49]
55	Spondyloepimetaphyseal dysplasia	AD	16p11.2	*KIF22*	19	HiSeq2000	SS	Boyden *et al.*[120]
					9	GAII	SS	Min *et al.*[121]
56	Adult neuronal ceroid-lipofuscinosis	AD	20q13.33	*DNAJC5*	1	SOLiD 4	SS	Nosková *et al.*[122]
57	KBG syndrome	AD	16q24.3	*ANKRD11*	3	GAIIx	NS	Sirmaci *et al.*[123]
58	Dendritic cell, monocyte, B and NK lymphoid deficiency	AD	3q21.3	*GATA-2*				Dickinson *et al.*[124]
59	Late-onset Parkinson's disease	AD	16q12	*VPS35*	2	GAIIx	SS	Zimprich *et al.*[125]
					2	GA	NA	Vilariño-Güell *et al.*[126]
60	Sensory neuropathy with dementia and hearing loss	AD	19p13.2	*DNMT1*	1	GAII/ 454 GS-FLX	SS/NA	Klein *et al.*[127]
61	Dilated cardiomyopathy	AD	10q25.2-q26.2	*BAG3*	4	GAIIx	NS	Norton *et al.*[128]
62	High myopia	AD	1p22.2	*ZNF644*	2	GAII	NA	Shi *et al.*[129]
63	Autosomal dominant retinitis pigmentosa	AD	1p31	*RPE65*				Bowne *et al.*[130]
64	Charcot–Marie–Tooth disease	AD	14q32	*DYNC1H1*	3	GAII	SS	Weedon *et al.*[131]
65	Hereditary hypotrichosis simplex	AD	13q12.2	*RPL21*	1	ND	NA	Zhou *et al.*[132]
66	Geleophysic and acromicric dysplasia	AD	15q21.1	*FBN1*	5	SOLiD 3.5	SS	Le Goff *et al.*a[133]
67	Myhre syndrome	AD	18q21.1	*SMAD4*	2	GAIIx	SS	Le Goff *et al.*[134]
68	Leucoencephalopathy	XLR	Xq13.2	*MCT8*	2	GAIIx	SS	Tsurusaki *et al.*[135]
2012								
69	Split hand and foot malformation	AR	7q22	*DLX5*	1	HiSeq2000	T	Shamseldin *et al.*[136]
70	Global eye developmental defects	AR	10q21.3-q22.1	*ATOH7*	1	GAIIx	SS	Khan *et al.*[137]
71	Primary hypertrophic osteoarthropathy	AR	3q21	*SLCO2A1*	1	GAII	NA	Zhang *et al.*[138]
72	Bartsocas–Papas Syndrome	AR	21q22.3	*RIPK4*	1	SOLiD	SS	Mitchell *et al.*[139]
73	Familial aplastic anemia	AR	1p34	*MPL*	1	GAIIx	NS	Walne *et al.*[140]
74	Peeling skin syndrome	AR	19q13.1	*CHST8*	1	SOLiD 4	SS	Cabral *et al.*[141]
75	Sengers syndrome	AR	7q34	*AGK*	1	GAIIx	SS	Mayr *et al.*[142]
76	Hypertension	AR/ AD AD	5q31 2q36.2	*KLHL3* *CUL3*	11	GAII	NA	Boyden *et al.*[143]
77	Weaver syndrome	AD	7q35-q36	*EZH2*	6	HiSeq2000	SS	Gibson *et al.*[144]
78	Genitopatellar syndrome	AD	10q22.2	*KAT6B*	6	GAIIx	SS	Simpson *et al.*[145]
					3	HiSeq	NS	Campeau *et al.*[146]
79	Hypothyroidism	AD	17q11.2	*THRA*	1	SOLiD4	SS	Bochukova *et al.*[147]
80	Floating–Harbor syndrome	AD	16p11.2	*SRCAP*	5	HiSeq	SS	Hood *et al.*[148]
81	Hereditary spastic paraplegia type 12	AD	19q13.32	*RTN2*	4	HiSeq2000	SS	Montenegro *et al.*[149]
82	Microcephaly associated with lymphedema	AD	10q24.1	*KIF11*	5	GAIIx	SS	Ostergaard *et al.*[150]
83	Congenital disorders of glycosylation (CDG)	AR	1p36.1	*DDOST*	1	HiSeq	NS	Jones *et al.*[151]
84	Congenital mirror movements	AD	15q15.1	*RAD51*	2	GAIIx	SS	Depienne *et al.*[48]

Table 2 (Continued)

Year	Disorder	MI	Location	Gene	Number exome sequenced	Sequencer	Capturing method	References
85	Mandibulofacialdysostosis with microcephaly	AD	17q21.31	*EFTUD2*	4	HiSeq	SS	Lines *et al.*[152]
86	Limb-girdle muscular dystrophy	AD	7q36.3	*DNAJB6*	3	HiSeq2000	T[e]	Harms *et al.*[153]
87	Congenital stationary night blindness	AR	17q21.1	*GPR179*	4	HiSeq	SS	Audo *et al.*[154]
88	Autosomal recessive primary microcephaly	AR	4q12	*CEP135*	1	HiSeq2000	N	Hussain *et al.*[155]
89	Aplastic anemia and myelodysplasia	AD	4q11	*SRP72*	4	GAIIx	N	Kirwan *et al.*[156]
90	Acrodysostosis	AD	5q12, 17q23-q24	*PDE4D, PRKAR1A*	5	HiSeq2000	SS	Lee *et al.*[157]
91	Olmsted syndrome	AD	17p13.2	*TRPV3*	13	HiSeq2000	SS	Lin *et al.*[158]
92	Familial diarrhea	AR	12p12	*GUCY2C*	3	HiSeq2000	NS	Fiskerstrand *et al.*[159]
93	Nager syndrome	AD	1q21.2	*SF3B4*	7	HiSeq2000	SS	Bernier *et al.*[160]
94	Infantile cerebellar retinal degeneration	AR	22q13.2	*ACO2*	1	SOLiD4	SS	Spiegel *et al.*[161]
95	Multicantric carpotarsal osteolysis	AD	20q11.2-q13.1	*MAFB*	5	GAII	NS	Zankl *et al.*[162]
96	Coffin–Siris syndrome	AD	22q11	*SMARCB1*	5	GAIIx	SS	Tsurusaki *et al.*[163]
97	Joubert syndrome	AR	5p13.2	*C5ORF42*	15	HiSeq2000	SS	Srour *et al.*[164]
98	Cerebroretinal microcephaly with calcifications and cysts	AR	17p13.1	*CTC1*	4	GAIIx	NS	Polvi *et al.*[165]
99	Kohlschutter–Tonz Syndrome	AR	16p13.3	*ROGDI*	1	GAII	SS	Schossig *et al.*[166]
100	Coffin–Siris syndrome	AD	6q25.1	*ARID1B*	5	HiSeq2000	SS	Santen *et al.*[167]
101	UV-sensitive syndrome	AR	4p16.3	*UVSSA*	2	GAIIx	SS	Nakazawa *et al.*[168]
102	Pulmonary arterial hypertension	AD	7q31.1	*CAV1*	4	SOLiD4	SS	Austin *et al.*[169]

Abbreviations: AD, autosomal dominant; ADK, adenylate kinase; AR, autosomal recessive; MI, mode of inheritance; UV, ultraviolet.
[a]GAII: Illumina Genome Analyzer II.
[b]S, solution-based capture method; SS, Agilent SureSelect Human All Exon Kit captures based on biotinylatedcRNA oligonucleotide baits; SA, Agilent SureSelect Array.
[c]ABI SOLiD system sequencing (Life Technologies, Carlsbad, CA, USA).
[d]N, NimbleGen; NA, NimbleGen Sequence Capture Human Exome 2.1M Array is based on array; NS, NimbleGenSeqCap EZ whole-exome is solution-based capture method.
[e]T, TruSeqExome Enrichment kit (Illumina).

Most pathogenic variants thus far identified are located in highly conserved regions of the genome.[18,19] It is believed that most of the functional variants are located in the coding exons.[20] Most (91.8%) of the functional variants of the protein-coding variations are due to nonsense/missense (~56%), small insertion/deletions (~24%), splicing (~10%) and regulatory (~1.8%) mutations (Human Gene Mutation Database professional 2011).[20] Overall, 85% of the disease-causing mutations are estimated to be located at protein-coding regions.[20,21] Accordingly, WES could elucidate at least 78% of causative variants.

BASICS OF WES

The approach for exome sequencing is based on probe-hybridization method to capture entire exons.[22,23] The whole process is categorized into three steps, namely sample preparation, hybridization and sequencing (Figure 1). Briefly, the first step is sample preparation, in which the genomic DNA is sheared by nebulization or sonication to get desired fragments of about 250 bp. The fragment ends are repaired by T4 DNA ligase. The process of 3′ A-tailing is performed followed by ligation of paired-end adaptor to the fragments. The final step for sample preparation is to amplify the prepared library for a few cycles. To enrich the prepared library, hybridization with a biotinylated oligo library (RNA baits for example, Agilant SureSelect (Santa Clara, CA, USA) (635 250 RNA probes of 120 bp) or DNA baits, for example, NimbleGen[22] (Madison, WI, USA) (2.1 million DNA probes of 60–90 bp)) is performed and captured by streptavidin beads. The quality and quantity of the exome library is analyzed by highly sensitive methods, such as Agilent 2100 Bioanalyzer before sequencing step. The exome library is sequenced in paired-end reads for example in Illumina (San Diego, CA, USA) to yield a 75–100 bases per read. Amplification of surface-bound individual fragments using an isothermal bridge amplification method produces clonal clusters of about 1000 identical molecules per cluster; one fragment is, therefore, attached to one surface oligonucleotide, endures cluster generation, and the replicate copies are sequenced to yield one sequence read. When DNA chain is growing, the first step of sequencing procedure is detecting the next added fluorescently labeled base (reversible terminator) by means of a sensitive device like charge-coupled device camera. The terminator is changed to a standard nucleotide by removing the dye. Repeating this cycle, sequentially, determines the next base. About 79% of the reported genes were determined using Illumina sequencing machines (Table 2).

After sequencing, the data is processed in three major steps, including mapping, variant calling and annotation steps (Figure 1). The sequence data is aligned with Burrows–Wheeler Aligner[24] tool against a reference sequence such as hg18/hg19 (GRCh37). Next step is calling; data generated by Burrows–Wheeler Aligner in Sequence Alignment Map (SAM) format could be used by SAMTools,[25] Genome Analysis Tool Kit,[26] and Picard (http://picard.sourceforge.net). SAMTools is used for quality control, short read alignment and variant identification (VarFilter). It processes and sorts the files. Facilitating the short aligned reads (BAM files: binary equivalent SAM format) for fast access is called indexing, which is followed by making a pileup format to facilitate variant calling. The

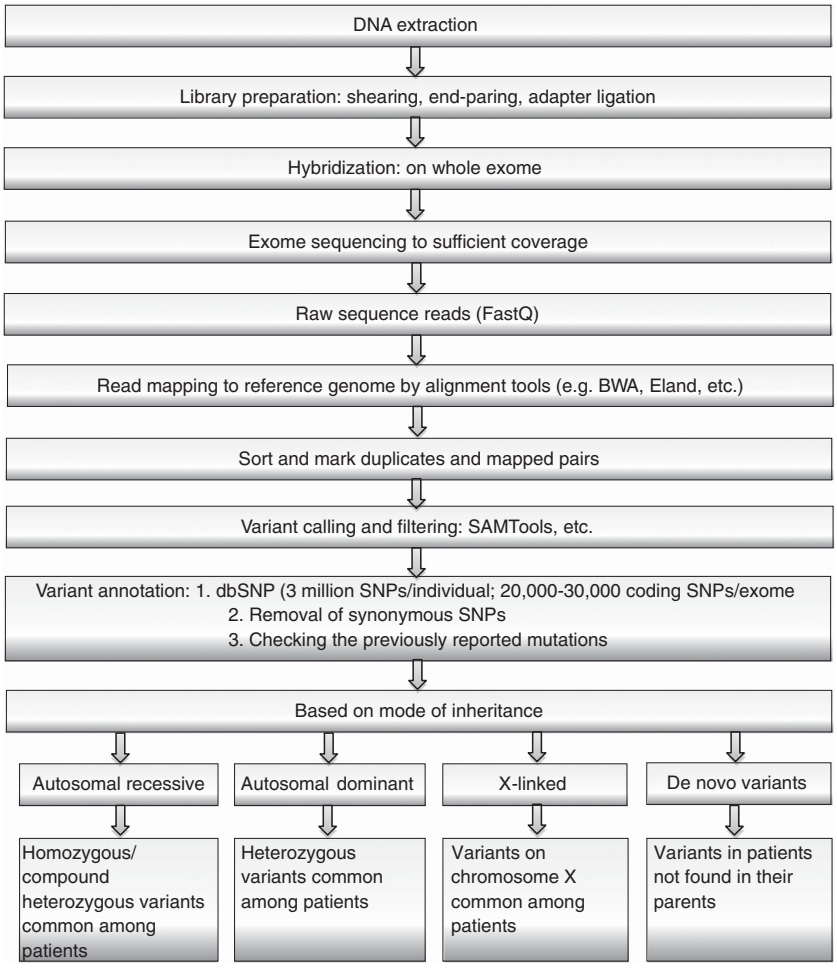

Figure 1 Applying usual filtering to exome-sequencing projects would define novel causal genes for Mendelian disorders; major assumptions about causal genes at these steps are as following: (1) structural variants and other forms of genetic variations are ignored, (2) causal variants are coding, (3) causal variants alter protein sequence and (4) casual variant has almost complete penetrance. A single exome carries about 20 000–30 000 coding SNPs. Over 95% of the variants overlap with data sets depending on ethnicity. Filtering steps narrow down the number of possible disease-associated genes; then, the final variants are limited to those fitting the mode of inheritance. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

indexed file is visualized by Integrative Genomics Viewer[27] or other sequence alignment visualization tools. PCR duplicates are removed using Picard MarkDuplicate and SAMTools. Average coverage and depth of coverage are calculated with Genome Analysis Tool Kit's Depth of Coverage analysis. ANNOVAR[28] is a tool for annotating genetic variants based on the function; the annotation file usually includes gene name, chromosomal position, nucleotide changes, amino-acid changes and description, SIFT (sorting intolerant from tolerant)[29] and Polyphen (polymorphism phenotyping)[30] values, single-nucleotide polymorphism database ID, allelic frequency of the SNP in 1000 Genome project and sequence quality. VAAST (Variant Annotation, Analysis and Search Tool) incorporated previous amino-acid substitution information with annotation and ranked candidate genes with statistical evaluation, which can be used to list up the candidate genes and variants.[31] Most investigators filter the data based on the function of variants. Nearly half of variants are synonymous ones, not considered to be deleterious, which are usually filtered out. Although there are some reports about the causal effect of synonymous variants,[32] the probability is very low. The remaining variants are nonsense, missense, indel, splice mutations and other non-coding RNA transcripts. Approximately, 5% of the variants are not reported in the above databases.[33] As noted, the variants called based on pathogenic predication of bioinformatics tools, such as SIFT[29] and PolyPhen,[30] are explored through the annotation. Hence, the pathogenic variants disrupt the protein function or structure in conserved sites. Depending on the knowledge of the affected samples, different analytic frameworks are used to define the causal variant (Figure 2).

CALLING VARIANTS AND THE CANDIDATE GENE

The sequence data are compared with public databases, such as single-nucleotide polymorphism database,[34] 1000 Genome Project,[35] Exome Variant Server (http://evs.gs.washington.edu/EVS) and HapMap.[36] It is noticed that individual exome of African-American origin has an average 24 000 single-nucleotide variants, whereas European-Americans origin has a mean 20 000 single-nucleotide variantss.[33] Thus, it is inferred from other studies that this number varies depending on the ethnicity and capturing protocols, sequencing platforms, mapping algorithms and variant calling methods. Totally, the number of candidate disease-causing variants is reduced to 100–500 pathogenic variants depending on the study design.[18,36–39] In a study, it was reported that each genome carries 165 homozygous

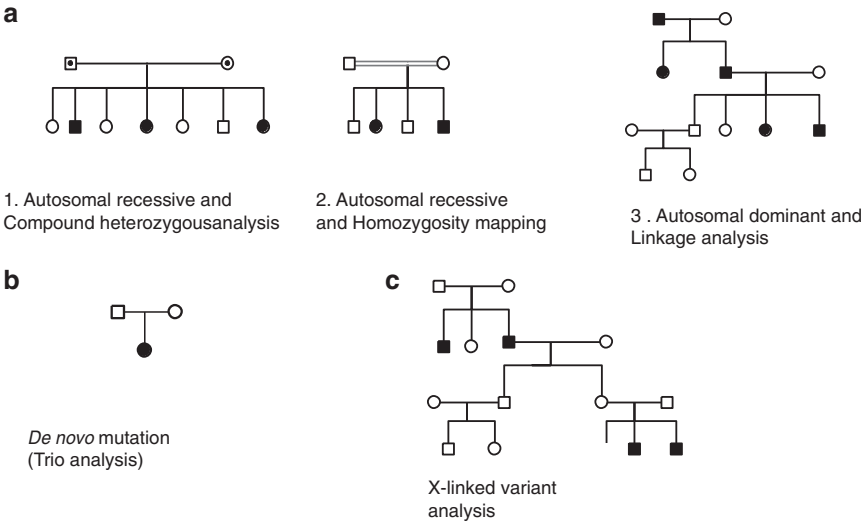

Figure 2 Hypothetical frameworks for analyzing single-gene disorders. Combinational analyses could help to determine the probable causative variant. Family-based (a1, a2 and a3), *de novo* mutation (**b**) and X-linked variant analysis (**c**). A full color version of this figure is available at the *Journal of Human Genetics* journal online.

protein-truncating variants in the diverse pathways.[40] Thus, a causal variant cannot be directly identified as the related gene unless integrative genomic analyses are performed as homozygosity mapping, linkage analysis, Sanger sequencing and so on. However, if we study a family with four affected individuals or two or three families, each with at least two affected ones, employing the usual filtering could possibly define a causal gene.[41–45] According to an assumption by Robinson *et al.*,[46] when the same gene is considered as a causality in multi sporadic cases, 5% of the target genes (about 20 000) show rare probable casual variants in all affected individuals, and after sequencing one individual and a usual filtering, nearly 1000 genes would remain as candidate genes. If a second individual is sequenced, only 50 genes (5% of 1000) with variants in both individuals will remain. Sequencing a third affected person predicts less than one gene having a variant in all three affected individuals.[46]

FRAMEWORKS USED TO ANALYZE SINGLE-GENE DISORDERS

We, here, exemplify the main approaches for detecting gene variants among the Mendelian disorders. Two main approaches, including family-based and unrelated individual strategies, are explained (Figure 2).

Family-based studies

When a number of affected individuals within a family are sequenced, the shared mutations are selected from the affected members because they harbor the same causal variant. This strategy narrow downs the candidate genes. Non-affected members of the family are sequenced to verify the candidate variations.

Combining the previous knowledge of homozygosity mapping for recessive disorders (Figure 2a2) and linkage studies for dominant (Figure 2a3) and recessive disorders as integrative approaches define the candidate variant. For instance, homozygous regions of the genome detected by SNP array are informative to reduce the number of candidate variants found to be homozygous for a family with recessive inheritance; only those variants in homozygous regions are reliable for pathogenicity. Shared homozygous or compound heterozygous variants are used to find the candidate variant (Figures 2a1 and a2). In a study by Sirmaci *et al.*,[47] the cause of Malpuech–Michels–Mingarelli–Carnevale syndrome in two affected families was identified. An autozygous region on chromosome 3q.27 was identified and exome sequencing confirmed *MASP1* mutation co-segregated with the phenotype.

Linkage studies are informative for multiple affected family members with multigenerations and are used in combination with exome sequencing.[48] For the dominant disorders, a common heterozygous variant is distinguished among the affected individuals in a family. Using the genome-wide linkage analysis of hereditary diffuse leukoencephalopathy with spheroids affecting central nervous system, Rademakers *et al.*[49] identified 233 candidate genes within the chromosome 5q candidate region and exome sequencing revealed a heterozygous variant in *CSF1R* in the candidate region, which was confirmed in 13 other affected families with distinct heterozygous mutations.

In case of locus heterogeneity in genetic disorders, such as retinitis pigmentosa, osteogenesis imperfect, hearing loss and so on, different patterns of inheritance may be observed in different families; thus, differentiating the exact clinical descriptions and determining the mode of inheritance would help to find the candidate gene. In a study by Abou Jamra *et al.*,[50] a combination of autozygosity mapping and exome sequencing was applied to identify the pathogenic variants, causing intellectual disability with recessive mode of inheritance in eight affected individuals from three consanguineous families. Using this approach, they identified three causative genes encoding adaptor protein complex 4 within these families.

In case of X-linked pedigrees (Figure 2c), analysis of X-chromosome variants could be helpful; of course, female and male samples are homozygous and hemizygous for autosomal recessive disorders, respectively. In an example, exome sequencing of entire three affected males having an unclassified X-linked lethal congenital malformation syndrome identified a splicing mutation in OFD1 gene.[51]

Unrelated individual studies

If there are a number of affected cases, but not within the family (sporadic cases), common pathogenic gene could be followed among the samples, assuming no locus heterogeneity among the affected individuals. The approach is called overlap strategy.[52] To point out,

the cause of Schinzel–Giedion syndrome was identified using this strategy.[53] Furthermore, Saitsu *et al.*[54] performed exome sequencing in three unrelated affected individuals with congenital hypomyelination leukoencephalopathy; they found compound heterozygous mutations in *POLR3A* and *POLR3B* (encoding RNA polymerase III subunits) in the affected individuals.

As in case of Kabuki syndrome, which is a rare disorder worldwide, the majority of cases are sporadic, but parent-to-child transmission has been reported representing the dominant mode of inheritance. WES revealed the causal variants of Kabuki syndrome in 7 out of 10 families in *MLL2*; follow-up Sanger sequencing of the remaining three families detected mutations in the *MLL2* in two of the families, which shows that this would be the main cause of syndrome.[39] Other genes, however, may explain the pathogenesis of the condition in the remaining family. One may use this family in conjunction with other new affected families to find the causal variants. In addition, they reported that only 26 out of 43 patients had mutations in *MLL2* by Sanger sequencing. The clinical and genetic heterogeneity of the syndrome complicates gene finding, and similar clinically affected cases are helpful to find new genes.

In the absence of number of cases, some integrative strategies are needed depending on their availability. When there is a single affected individual suspected of a recessive disorder, homozygous and compound heterozygous variants are searched namely, as double hit strategy by Gilissen *et al.*[52] WES revealed compound heterozygous mutations in *WDR35* gene in a single sporadic case affected with Sensenbrenner syndrome.[55] Also, WES of one of the affected sisters with Perrault syndrome manifested two mutations, showing compound heterozygote in the *HSD17B4*.[38]

De novo mutations

The previously mentioned frameworks are focused on homogenous diseases. A substantial number of *de novo* mutations occur sporadically in which cases, mostly fetus, do not survive and so the mutations will be eliminated from the population; thus, these lethal mutations are not usually identified. *De novo* mutation rate is estimated to be 7.6×10^{-9}–2.2×10^{-8} per generation; that is, approximately one in 10^8 base per haploid genome is mutated spontaneously,[56,57] which could become the causality. For the *de novo* mutation analysis, the case–parent trios are practical (Figure 2b). Nonpathogenic variants are filtered and then the variants presented in the parents are excluded. There might be a chance of sequencing errors and mapping artefacts, so confirmation by Sanger sequencing with high accuracy should be applied.[58] Detection of a *de novo* mutation is not enough to confirm the causality of the disease. Additional analyses, including replication and functional analyses, should be performed to determine the deleterious or causal variants. Pathogenesity of a variant not only depends on the type and the location of the mutation but also on its functional effects.[19]

De novo mutation studies have been employed to determine *de novo* mutations of rare Mendelian disorders, such as Schinzel–Giedion syndrome.[53] In addition, causative genetic factors for heterogeneous disorders, such as intellectual disabilities, have been revealed as well.[18] Trio-based exome sequencing is demonstrated to be a powerful approach for identifying novel causative genes for sporadic autism spectrum;[59] these researchers identified 21 *de novo* mutations using exome sequencing of 20 sporadic cases of the sporadic autism spectrum.[59] More recently, Sanders *et al.*[60] demonstrated, using WES of families affected by autism spectrum disorder, contribution of *de novo* mutations in brain-expression genes to the risk for these disorders. Iossifov *et al.*[61] sequenced and analyzed the exomes of 343 families with a single individual affected by the autism spectrum and at least one unaffected sibling; they found that gene-disrupting mutations, not missense mutations, are frequent in affected children. In this study, 350–400 genes have been estimated as autism-susceptibility genes.[61]

PROS AND CONS

Interpretation of the results obtained by NGS of the Mendelian disorders is of major concern. When using exome sequencing in clinical genetics and medicine, limitation of the approach is evident and experimental design is needed to circumvent the downs. Genetic and phenotypic heterogeneity in different affected individuals make exome sequencing difficult to interpret. Exact clinical examinations and biochemical tests have important roles to distinguish between new syndromes and known ones.

Patients with the same phenotype may not share the same causal variant; indeed they may have distinct variants in a gene as we call allelic heterogeneity. Depending on the clinical information, different strategies or filtering protocols are used for implication of the pathogenicity of a variant. Of course, variants in the reported genes are generally examined at the first step of filtering process. Then, the possible causal variants could be validated by segregating through the family or other cases.

Intensive examinations of variant call are important in exome sequencing; false-positive errors appear as sequencing errors related to mechanical and analytical errors; also short reads generated by NGS would not align perfectly to the appropriate position as a result of paralogous and low copy repeats that may cause errors during calling.[62] In the repetitive regions of the genome, misalignment may occur, which could be improved by longer read lengths or higher depth of coverage in those regions.[21] Also, false-negative errors could occur because of mechanical and analytic errors due to low coverage, poor capture efficiency and so on. Avoiding false-negative errors is more difficult than false-positive ones; however, it is proposed that the error rate could be estimated by comparing the calls with the test samples, which have been previously called.[21,46] The relation of false-positive and false-negative calls is 'trade-off'; if we set stringent criteria (high base quality and stringent alignment or and so on), false-positive errors are decreased but false-negative ones are increased, which is important to concern. There are also other problems in the filtering strategies, which may influence data analysis. Filtering out variants with minor allele frequency of $<1\%$ may be misleading for recessive disorders. Because the carriers may not show the disease, but still the frequency of the allele is high in the 1000 Genome, which may be excluded in the filtering step because of higher frequency.

Some deleterious variations may be located in the non-coding regions, such as intronic or regulatory regions of the genome, which cannot be called by exome sequencing, whereas WGS covers all the data for genome. WGS is expected to be applied for disease-gene identification in near future; however, the current cost and information burden of WGS need to be circumvented. The huge amount of data generated from WGS comparing with WES provides information on evolutionary-conserved non-coding regions and all variants throughout the genome. Filtering and analyzing these data is challenging. Moreover, the time for analyzing data is increased and larger computational memory is needed for WGS data analysis.

CONCLUSION

Exome sequencing has evolved the biomedical research. The possible causative genes are directly distinguished using these new sequencing

technologies. Up to now, the role of more than 100 genes has been distinguished in rare Mendelian disorders by means of WES, and this statistics is rapidly growing. Combinational approaches, including traditional methods and WES, are easily used for those disorders following autosomal mode of inheritance to define the underlying gene. It is needless to say, new sequencing technologies, such as in Pacific Bioscience and Nanopore, will shed light in this field.[63,64] WES could have a critical role in identifying new genes until the costs for WGS will be decreased.

ACKNOWLEDGEMENTS

We would like to thank the staff of Division of Human Genetics at National Institute of Genetics, Mishima, Shizuoka, Japan.

1 Green, E. D. & Guyer, M. S. Charting a course for genomic medicine from base pairs to bedside. *Nature* **470,** 204–213 (2011).
2 Reich, D. E. & Lander, E. S. On the allelic spectrum of human disease. *Trends Genet.* **17,** 502–510 (2001).
3 Lander, E. S. Initial impact of the sequencing of the human genome. *Nature* **470,** 187–197 (2011).
4 Pritchard, J. K. Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.* **69,** 124–137 (2001).
5 Gorlov, I. P., Gorlova, O. Y., Sunyaev, S. R., Spitz, M. R. & Amos, C. I. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am. J. Hum. Genet.* **82,** 100–112 (2008).
6 McKusick, V. A. Mendelian Inheritance in Man and its online version, OMIM. *Am. J. Hum. Genet.* **80,** 588–604 (2007).
7 Antonarakis, S. E. & Beckmann, J. S. Mendelian disorders deserve more attention. *Nat. Rev. Genet.* **7,** 277–282 (2006).
8 Peltonen, L., Perola, M., Naukkarinen, J. & Palotie, A. Lessons from studying monogenic disease for common disease. *Hum Mol Genet* **15(Spec No 1),** R67–R74.
9 Holley, R. W. The nucleotide sequence of a nucleic acid. *Sci. Am.* **214,** 30–39 (1966).
10 Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* **74,** 5463–5467 (1977).
11 Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. *Proc. Natl Acad. Sci. USA* **74,** 560–564 (1977).
12 Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* **42,** 30–35 (2010).
13 Lindhurst, M. J., Sapp, J. C., Teer, J. K., Johnston, J. J., Finn, E. M., Peters, K. *et al.* A mosaic activating mutation in AKT1 associated with the Proteus syndrome. *N. Engl. J. Med.* **365,** 611–619 (2011).
14 Riordan, J. R., Rommens, J. M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z. *et al.* Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **245,** 1066–1073 (1989).
15 Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E. *et al.* A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306,** 234–238 (1983).
16 Lander, E. S. & Botstein, D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* **236,** 1567–1570 (1987).
17 Pomares, E., Riera, M., Permanyer, J., Mendez, P., Castro-Navarro, J., Andres-Gutierrez, A. *et al.* Comprehensive SNP-chip for retinitis pigmentosa-leber congenital amaurosis diagnosis: new mutations and detection of mutational founder effects. *Eur. J. Hum. Genet.* **18,** 118–124 (2010).
18 Vissers, L. E., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P. *et al.* A *de novo* paradigm for mental retardation. *Nat. Genet.* **42,** 1109–1112 (2010).
19 Cooper, D. N., Chen, J. M., Ball, E. V., Howells, K., Mort, M., Phillips, A. D. *et al.* Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics. *Hum. Mutat.* **31,** 631–655 (2011).
20 Botstein, D. & Risch, N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.* **33(Suppl),** 228–237 (2003).
21 Majewski, J., Schwartzentruber, J., Lalonde, E., Montpetit, A. & Jabado, N. What can exome sequencing do for you? *J. Med. Genet.* **48,** 580–589 (2011).
22 Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M. N., Smith, S. W. *et al.* Genome-wide *in situ* exon capture for selective resequencing. *Nat. Genet.* **39,** 1522–1527 (2007).
23 Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* **27,** 182–189 (2009).
24 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
25 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
26 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).
27 Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29,** 24–26 (2011).
28 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164 (2010).
29 Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4,** 1073–1081 (2009).
30 Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7,** 248–249 (2010).
31 Yandell, M., Huff, C., Hu, H., Singleton, M., Moore, B., Xing, J., Jorde, L. B. *et al.* A probabilistic disease-gene finder for personal genomes. *Genome Res.* **21,** 1529–1542 (2011).
32 Sauna, Z. E. & Kimchi-Sarfaty, C. Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet.* **12,** 683–691 (2011).
33 Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* **12,** 745–755 (2011).
34 Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29,** 308–311 (2001).
35 Consortium, G. P. A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).
36 Altshuler, D. M., Gibbs, R. A., Peltonen, L., Dermitzakis, E., Schaffner, S. F., Yu, F. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467,** 52–58 (2010).
37 Musunuru, K., Pirruccello, J. P., Do, R., Peloso, G. M., Guiducci, C., Sougnez, C. *et al.* Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N. Engl. J. Med.* **363,** 2220–2227 (2010).
38 Pierce, S. B., Walsh, T., Chisholm, K. M., Lee, M. K., Thornton, A. M., Fiumara, A. *et al.* Mutations in the DBP-deficiency protein HSD17B4 cause ovarian dysgenesis, hearing loss, and ataxia of Perrault syndrome. *Am. J. Hum. Genet.* **87,** 282–288 (2010).
39 Ng, S. B., Bigham, A. W., Buckingham, K. J., Hannibal, M. C., McMillin, M. J., Gildersleeve, H. I. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat. Genet.* **42,** 790–793 (2010).
40 Pelak, K., Shianna, K. V., Ge, D., Maia, J. M., Zhu, M., Smith, J. P. *et al.* The characterization of twenty sequenced human genomes. *PLoS Genet* **6,** e1001111 (2010).
41 Zangen, D., Kaufman, Y., Zeligson, S., Perlberg, S., Fridman, H., Kanaan, M. *et al.* XX ovarian dysgenesis is caused by a PSMC3IP/HOP2 mutation that abolishes coactivation of estrogen-driven transcription. *Am. J. Hum. Genet.* **89,** 572–579 (2011).
42 Pierson, T. M., Adams, D., Bonn, F., Martinelli, P., Cherukuri, P. F., Teer, J. K. *et al.* Whole-exome sequencing identifies homozygous AFG3L2 mutations in a spastic ataxia-neuropathy syndrome linked to mitochondrial m-AAA proteases. *PLoS Genet* **7,** e1002325 (2011).
43 Puente, X. S., Quesada, V., Osorio, F. G., Cabanillas, R., Cadinanos, J., Fraile, J. M. *et al.* Exome sequencing and functional analysis identifies BANF1 mutation as the cause of a hereditary progeroid syndrome. *Am. J. Hum. Genet.* **88,** 650–656 (2011).
44 Wang, J. L., Yang, X., Xia, K., Hu, Z. M., Weng, L., Jin, X. *et al.* TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. *Brain* **133(Part 12),** 3510–3518 (2010).
45 Sun, Y., Almomani, R., Aten, E., Celli, J., van der Heijden, J., Venselaar, H. *et al.* Terminal osseous dysplasia is caused by a single recurrent mutation in the FLNA gene. *Am. J. Hum. Genet.* **87,** 146–153 (2010).
46 Robinson, P. N., Krawitz, P. & Mundlos, S. Strategies for exome and genome sequence data analysis in disease-gene discovery projects. *Clin. Genet.* **80,** 127–132 (2011).
47 Sirmaci, A., Walsh, T., Akay, H., Spiliopoulos, M., Sakalar, Y. B., Hasanefendioglu-Bayrak, A. *et al.* MASP1 mutations in patients with facial, umbilical, coccygeal, and auditory findings of Carnevale, Malpuech, OSA, and Michels syndromes. *Am. J. Hum. Genet.* **87,** 679–686 (2010).
48 Depienne, C., Bouteiller, D., Meneret, A., Billot, S., Groppa, S., Klebe, S. *et al.* RAD51 haploinsufficiency causes congenital mirror movements in humans. *Am. J. Hum. Genet.* **90,** 301–307 (2012).
49 Rademakers, R., Baker, M., Nicholson, A. M., Rutherford, N. J., Finch, N., Soto-Ortolaza, A. *et al.* Mutations in the colony stimulating factor 1 receptor (CSF1R) gene cause hereditary diffuse leukoencephalopathy with spheroids. *Nat. Genet.* **44,** 200–205 (2011).
50 Abou Jamra, R., Philippe, O., Raas-Rothschild, A., Eck, S. H., Graf, E., Buchert, R. *et al.* Adaptor protein complex 4 deficiency causes severe autosomal-recessive intellectual disability, progressive spastic paraplegia, shy character, and short stature. *Am. J. Hum. Genet.* **88,** 788–795 (2011).
51 Tsurusaki, Y., Kosho, T., Hatasaki, K., Narumi, Y., Wakui, K., Fukushima, Y. *et al.* Exome sequencing in a family with an X-linked lethal malformationsyndrome: clinical consequences of hemizygous truncating OFD1 mutations in male patients. *Clin. Genet, doi:10.1111/j.1399-0004.2012.01885.x (2012).*
52 Gilissen, C., Hoischen, A., Brunner, H. G. & Veltman, J. A. Disease gene identification strategies for exome sequencing. *Eur. J. Hum. Genet.* **20,** 490–497 (2011).

53 Hoischen, A., van Bon, B. W., Gilissen, C., Arts, P., van Lier, B., Steehouwer, M. *et al.* *De novo* mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat. Genet.* **42,** 483–485 (2010).
54 Saitsu, H., Osaka, H., Sasaki, M., Takanashi, J., Hamada, K., Yamashita, A. *et al.* Mutations in POLR3A and POLR3B encoding RNA Polymerase III subunits cause an autosomal-recessive hypomyelinating leukoencephalopathy. *Am. J. Hum. Genet.* **89,** 644–651 (2011).
55 Gilissen, C., Arts, H. H., Hoischen, A., Spruijt, L., Mans, D. A., Arts, P. *et al.* Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome. *Am. J. Hum. Genet.* **87,** 418–423 (2010).
56 Roach, J. C., Glusman, G., Smit, A. F., Huff, C. D., Hubley, R., Shannon, P. T. *et al.* Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328,** 636–639 (2010).
57 Lynch, M. Rate molecular spectrum, and consequences of human mutation. *Proc. Natl Acad. Sci. USA* **107,** 961–968 (2010).
58 Gilissen, C., Hoischen, A., Brunner, H. G. & Veltman, J. A. Unlocking Mendelian disease using exome sequencing. *Genome Biol.* **12,** 228 (2011).
59 O'Roak, B. J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J. J., Girirajan, S. *et al.* Exome sequencing in sporadic autism spectrum disorders identifies severe *de novo* mutations. *Nat. Genet.* **43,** 585–589 (2011).
60 Sanders, S. J., Murtha, M. T., Gupta, A. R., Murdoch, J. D., Raubeson, M. J., Willsey, A. J. *et al. De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485,** 237–241 (2012).
61 Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J. *et al. De novo* gene disruptions in children on the autistic spectrum. *Neuron* **74,** 285–299 (2012).
62 Koboldt, D. C., Ding, L., Mardis, E. R. & Wilson, R. K. Challenges of sequencing human genomes. *Brief Bioinform.* **11,** 484–498 (2010).
63 Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323,** 133–138 (2009).
64 Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T. *et al.* The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* **26,** 1146–1153 (2008).
65 Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171,** 737–738 (1953).
66 Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32,** 314–331 (1980).
67 Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* **51,** 263–273 (1986).
68 Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R. *et al.* Fluorescence detection in automated DNA sequence analysis. *Nature* **321,** 674–679 (1986).
69 Royer-Pokora, B., Kunkel, L. M., Monaco, A. P., Goff, S. C., Newburger, P. E., Baehner, R. L. *et al.* Cloning the gene for an inherited human disorder—chronic granulomatous disease—on the basis of its chromosomal location. *Nature* **322,** 32–38 (1986).
70 Cohen, D., Chumakov, I. & Weissenbach, J. A first-generation physical map of the human genome. *Nature* **366,** 698–701 (1993).
71 Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269,** 496–512 (1995).
72 Dunham, I., Shimizu, N., Roe, B. A., Chissoe, S., Hunt, A. R., Collins, J. E. *et al.* The DNA sequence of human chromosome 22. *Nature* **402,** 489–495 (1999).
73 Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G. *et al.* The genome sequence of *Drosophila melanogaster. Science* **287,** 2185–2195 (2000).
74 Myers, E. W., Sutton, G. G., Smith, H. O., Adams, M. D. & Venter, J. C. On the sequencing and assembly of the human genome. *Proc. Natl Acad. Sci. USA* **99,** 4145–4146 (2002).
75 Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G. *et al.* The sequence of the human genome. *Science* **291,** 1304–1351 (2001).
76 Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
77 Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437,** 376–380 (2005).
78 Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452,** 872–876 (2008).
79 Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461,** 272–276 (2009).
80 Bolze, A., Byun, M., McDonald, D., Morgan, N. V., Abhyankar, A., Premkumar, L. *et al.* Whole-exome-sequencing-based discovery of human FADD deficiency. *Am. J. Hum. Genet.* **87,** 873–881 (2010).
81 Walsh, T., Shahin, H., Elkan-Miller, T., Lee, M. K., Thornton, A. M., Roeb, W. *et al.* Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82. *Am. J. Hum. Genet.* **87,** 90–94 (2010).
82 Haack, T. B., Danhauser, K., Haberberger, B., Hoser, J., Strecker, V., Boehm, D. *et al.* Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency. *Nat. Genet.* **42,** 1131–1134 (2010).
83 Krawitz, P. M., Schweiger, M. R., Rodelsperger, C., Marcelis, C., Kolsch, U., Meisel, C. *et al.* Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome. *Nat. Genet.* **42,** 827–829 (2010).
84 Bilguvar, K., Ozturk, A. K., Louvi, A., Kwan, K. Y., Choi, M., Tatli, B. *et al.* Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature* **467,** 207–210 (2010).
85 Caliskan, M., Chong, J. X., Uricchio, L., Anderson, R., Chen, P., Sougnez, C. *et al.* Exome sequencing reveals a novel mutation for autosomal recessive non-syndromic mental retardation in the TECR gene on chromosome 19p13. *Hum. Mol. Genet.* **20,** 1285–1289 (2011).
86 Zuchner, S., Dallman, J., Wen, R., Beecham, G., Naj, A., Farooq, A. *et al.* Whole-exome sequencing links a variant in DHDDS to retinitis pigmentosa. *Am. J. Hum. Genet.* **88,** 201–206 (2011).
87 Becker, J., Semler, O., Gilissen, C., Li, Y., Bolz, H. J., Giunta, C. *et al.* Exome sequencing identifies truncating mutations in human SERPINF1 in autosomal-recessive osteogenesis imperfecta. *Am. J. Hum. Genet.* **88,** 362–371 (2011).
88 Glazov, E. A., Zankl, A., Donskoi, M., Kenna, T. J., Thomas, G. P., Clark, G. R. *et al.* Whole-exome re-sequencing in a family quartet identifies POP1 mutations as the cause of a novel skeletal dysplasia. *PLoS Genet.* **7,** e1002027 (2011).
89 Sloan, J. L., Johnston, J. J., Manoli, I., Chandler, R. J., Krause, C., Carrillo-Carrasco, N. *et al.* Exome sequencing identifies ACSF3 as a cause of combined malonic and methylmalonic aciduria. *Nat. Genet.* **43,** 883–886 (2011).
90 O'Sullivan, J., Bitu, C. C., Daly, S. B., Urquhart, J. E., Barron, M. J., Bhaskar, S. S. *et al.* Whole-exome sequencing identifies FAM20A mutations as a cause of amelogenesis imperfecta and gingival hyperplasia syndrome. *Am. J. Hum. Genet.* **88,** 616–620 (2011).
91 Vissers, L. E., Lausch, E., Unger, S., Campos-Xavier, A. B., Gilissen, C., Rossi, A. *et al.* Chondrodysplasia and abnormal joint development associated with mutations in IMPAD1, encoding the Golgi-resident nucleotide phosphatase, gPAPP. *Am. J. Hum. Genet.* **88,** 608–615 (2011).
92 Gotz, A., Tyynismaa, H., Euro, L., Ellonen, P., Hyotylainen, T., Ojala, T. *et al.* Exome sequencing identifies mitochondrial alanyl-tRNA synthetase mutations in infantile mitochondrial cardiomyopathy. *Am. J. Hum. Genet.* **88,** 635–642 (2011).
93 Tariq, M., Belmont, J. W., Lalani, S., Smolarek, T. & Ware, S. M. SHROOM3 is a novel candidate for heterotaxy identified by whole exome sequencing. *Genome Biol.* **12,** R91 (2011).
94 Snape, K., Hanks, S., Ruark, E., Barros-Nunez, P., Elliott, A., Murray, A. *et al.* Mutations in CEP57 cause mosaic variegated aneuploidy syndrome. *Nat. Genet.* **43,** 527–529 (2011).
95 Theis, J. L., Sharpe, K. M., Matsumoto, M. E., Chai, H. S., Nair, A. A., Theis, J. D. *et al.* Homozygosity mapping and exome sequencing reveal GATAD1 mutation in autosomal recessive dilated cardiomyopathy. *Circ. Cardiovasc. Genet.* **4,** 585–594 (2011).
96 Takata, A., Kato, M., Nakamura, M., Yoshikawa, T., Kanba, S., Sano, A. *et al.* Exome sequencing identifies a novel missense variant in RRM2B associated with autosomal recessive progressive external ophthalmoplegia. *Genome Biol.* **12,** R92 (2011).
97 Aldahmesh, M. A., Khan, A. O., Mohamed, J. Y., Alkuraya, H., Ahmed, H., Bobis, S. *et al.* Identification of ADAMTS18 as a gene mutated in Knobloch syndrome. *J. Med. Genet.* **48,** 597–601 (2011).
98 Doi, H., Yoshida, K., Yasuda, T., Fukuda, M., Fukuda, Y., Morita, H. *et al.* Exome sequencing reveals a homozygous SYT14 mutation in adult-onset, autosomal-recessive spinocerebellar ataxia with psychomotor retardation. *Am. J. Hum. Genet.* **89,** 320–327 (2011).
99 Shaheen, R., Faqeih, E., Sunker, A., Morsy, H., Al-Sheddi, T., Shamseldin, H. E. *et al.* Recessive mutations in DOCK6, encoding the guanidine nucleotide exchange factor DOCK6, lead to abnormal actin cytoskeleton organization and Adams-Oliver syndrome. *Am. J. Hum. Genet.* **89,** 328–333 (2011).
100 Sanna-Cherchi, S., Burgess, K. E., Nees, S. N., Caridi, G., Weng, P. L., Dagnino, M. *et al.* Exome sequencing identified MYO1E and NEIL1 as candidate genes for human autosomal recessive steroid-resistant nephrotic syndrome. *Kidney Int.* **80,** 389–396 (2011).
101 Barak, T., Kwan, K. Y., Louvi, A., Demirbilek, V., Saygi, S., Tuysuz, B. *et al.* Recessive LAMC3 mutations cause malformations of occipital cortical development. *Nat. Genet.* **43,** 590–594 (2011).
102 Galmiche, L., Serre, V., Beinat, M., Assouline, Z., Lebre, A. S., Chretien, D. *et al.* Exome sequencing identifies MRPL3 mutation in mitochondrial cardiomyopathy. *Hum. Mutat.* **32,** 1225–1231 (2011).
103 Ozgul, R. K., Siemiatkowska, A. M., Yucel, D., Myers, C. A., Collin, R. W., Zonneveld, M. N. *et al.* Exome sequencing and cis-regulatory mapping identify mutations in MAK, a gene encoding a regulator of ciliary length, as a cause of retinitis pigmentosa. *Am. J. Hum. Genet.* **89,** 253–264 (2011).
104 Hanson, D., Murray, P. G., O'Sullivan, J., Urquhart, J., Daly, S., Bhaskar, S. S. *et al.* Exome sequencing identifies CCDC8 mutations in 3-M syndrome, suggesting that CCDC8 contributes in a pathway with CUL7 and OBSL1 to control human growth. *Am. J. Hum. Genet.* **89,** 148–153 (2011).
105 Kalay, E., Yigit, G., Aslan, Y., Brown, K. E., Pohl, E., Bicknell, L. S. *et al.* CEP152 is a genome maintenance protein disrupted in Seckel syndrome. *Nat. Genet.* **43,** 23–26 (2010).
106 Bjursell, M. K., Blom, H. J., Cayuela, J. A., Engvall, M. L., Lesko, N., Balasubramaniam, S. *et al.* Adenosine kinase deficiency disrupts the methionine cycle and causes hypermethioninemia, encephalopathy, and abnormal liver function. *Am. J. Hum. Genet.* **89,** 507–515 (2011).

107 Bredrup, C., Saunier, S., Oud, M. M., Fiskerstrand, T., Hoischen, A., Brackman, D. *et al*. Ciliopathies with skeletal anomalies and renal insufficiency due to mutations in the IFT-A gene WDR19. *Am. J. Hum. Genet.* **89,** 634–643 (2011).
108 Aldahmesh, M. A., Mohamed, J. Y., Alkuraya, H. S., Verma, I. C., Puri, R. D., Alaiya, A. A. *et al.* Recessive mutations in ELOVL4 cause ichthyosis, intellectual disability, and spastic quadriplegia. *Am. J. Hum. Genet.* **89,** 745–750 (2011).
109 Dauber, A., Nguyen, T. T., Sochett, E., Cole, D. E., Horst, R., Abrams, S. A. *et al.* Genetic defect in CYP24A1, the vitamin D 24-hydroxylase gene, in a patient with severe infantile hypercalcemia. *J. Clin. Endocrinol. Metab.* **97,** E268–E274 (2011).
110 Logan, C. V., Lucke, B., Pottinger, C., Abdelhamed, Z. A., Parry, D. A., Szymanska, K. *et al*. Mutations in MEGF10, a regulator of satellite cell myogenesis, cause early onset myopathy, areflexia, respiratory distress and dysphagia (EMARDD). *Nat. Genet.* **43,** 1189–1192 (2011).
111 Albers, C. A., Cvejic, A., Favier, R., Bouwmans, E. E., Alessi, M. C., Bertone, P. *et al.* Exome sequencing identifies NBEAL2 as the causative gene for gray platelet syndrome. *Nat. Genet.* **43,** 735–737 (2011).
112 de Greef, J. C., Wang, J., Balog, J., den Dunnen, J. T., Frants, R. R., Straasheijm, K. R. *et al.* Mutations in ZBTB24 are associated with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. *Am. J. Hum. Genet.* **88,** 796–804 (2011).
113 Sergouniotis, P. I., Davidson, A. E., Mackay, D. S., Li, Z., Yang, X., Plagnol, V. *et al.* Recessive mutations in KCNJ13, encoding an inwardly rectifying potassium channel subunit, cause leber congenital amaurosis. *Am. J. Hum. Genet.* **89,** 183–190 (2011).
114 Erlich, Y., Edvardson, S., Hodges, E., Zenvirt, S., Thekkat, P., Shaag, A. *et al.* Exome sequencing and disease-network analysis of a single family implicate a mutation in KIF1A in hereditary spastic paraparesis. *Genome Res.* **21,** 658–664 (2011).
115 Clayton-Smith, J., O'Sullivan, J., Daly, S., Bhaskar, S., Day, R., Anderson, B. *et al.* Whole-exome-sequencing identifies mutations in histone acetyltransferase gene KAT6B in individuals with the Say-Barber-Biesecker variant of Ohdo syndrome. *Am. J. Hum. Genet.* **89,** 675–681 (2011).
116 Chen, W. J., Lin, Y., Xiong, Z. Q., Wei, W., Ni, W., Tan, G. H. *et al.* Exome sequencing identifies truncating mutations in PRRT2 that cause paroxysmal kinesigenic dyskinesia. *Nat. Genet.* **43,** 1252–1255 (2011).
117 Simpson, M. A., Irving, M. D., Asilmaz, E., Gray, M. J., Dafou, D., Elmslie, F. V. *et al.* Mutations in NOTCH2 cause Hajdu-Cheney syndrome, a disorder of severe and progressive bone loss. *Nat. Genet.* **43,** 303–305 (2011).
118 Isidor, B., Lindenbaum, P., Pichon, O., Bezieau, S., Dina, C., Jacquemont, S. *et al.* Truncating mutations in the last exon of NOTCH2 cause a rare skeletal disorder with osteoporosis. *Nat. Genet.* **43,** 306–308 (2011).
119 Hoischen, A., van Bon, B. W., Rodriguez-Santiago, B., Gilissen, C., Vissers, L. E., de Vries, P. *et al. De novo* nonsense mutations in ASXL1 cause Bohring–Opitz syndrome. *Nat. Genet.* **43,** 729–731 (2011).
120 Boyden, E. D., Campos-Xavier, A. B., Kalamajski, S., Cameron, T. L., Suarez, P., Tanackovich, G. *et al.* Recurrent dominant mutations affecting two adjacent residues in the motor domain of the monomeric kinesin KIF22 result in skeletal dysplasia and joint laxity. *Am. J. Hum. Genet.* **89,** 767–772 (2011).
121 Min, B. J., Kim, N., Chung, T., Kim, O. H., Nishimura, G., Chung, C. Y. *et al.* Whole-exome sequencing identifies mutations of KIF22 in spondyloepimetaphyseal dysplasia with joint laxity, leptodactylic type. *Am. J. Hum. Genet.* **89,** 760–766 (2011).
122 Noskova, L., Stranecky, V., Hartmannova, H., Pristoupilova, A., Baresova, V., Ivanek, R. *et al.* Mutations in DNAJC5, encoding cysteine-string protein alpha, cause autosomal-dominant adult-onset neuronal ceroid lipofuscinosis. *Am. J. Hum. Genet.* **89,** 241–252 (2011).
123 Sirmaci, A., Spiliopoulos, M., Brancati, F., Powell, E., Duman, D., Abrams, A. *et al.* Mutations in ANKRD11 cause KBG syndrome, characterized by intellectual disability, skeletal malformations, and macrodontia. *Am. J. Hum. Genet.* **89,** 289–294 (2011).
124 Dickinson, R. E., Griffin, H., Bigley, V., Reynard, L. N., Hussain, R., Haniffa, M. *et al.* Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. *Blood* **118,** 2656–2658 (2011).
125 Zimprich, A., Benet-Pages, A., Struhal, W., Graf, E., Eck, S. H., Offman, M. N. *et al.* A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. *Am. J. Hum. Genet.* **89,** 168–175 (2011).
126 Vilarino-Guell, C., Wider, C., Ross, O. A., Dachsel, J. C., Kachergus, J. M., Lincoln, S. J. *et al.* VPS35 mutations in Parkinson disease. *Am. J. Hum. Genet.* **89,** 162–167 (2011).
127 Klein, C. J., Botuyan, M. V., Wu, Y., Ward, C. J., Nicholson, G. A., Hammans, S. *et al.* Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. *Nat. Genet.* **43,** 595–600 (2011).
128 Norton, N., Li, D., Rieder, M. J., Siegfried, J. D., Rampersaud, E., Zuchner, S. *et al.* Genome-wide studies of copy number variation and exome sequencing identify rare variants in BAG3 as a cause of dilated cardiomyopathy. *Am. J. Hum. Genet.* **88,** 273–282 (2011).
129 Shi, Y., Li, Y., Zhang, D., Zhang, H., Lu, F., Liu, X. *et al.* Exome sequencing identifies ZNF644 mutations in high myopia. *PLoS Genet.* **7,** e1002084 (2011).
130 Bowne, S. J., Humphries, M. M., Sullivan, L. S., Kenna, P. F., Tam, L. C., Kiang, A. S. *et al.* A dominant mutation in RPE65 identified by whole-exome sequencing causes retinitis pigmentosa with choroidal involvement. *Eur. J. Hum. Genet.* **19,** 1074–1081 (2011).
131 Weedon, M. N., Hastings, R., Caswell, R., Xie, W., Paszkiewicz, K., Antoniadi, T. *et al.* Exome sequencing identifies a DYNC1H1 mutation in a large pedigree with dominant axonal Charcot-Marie-Tooth disease. *Am. J. Hum. Genet.* **89,** 308–312 (2011).
132 Zhou, C., Zang, D., Jin, Y., Wu, H., Liu, Z., Du, J. *et al.* Mutation in ribosomal protein L21 underlies hereditary hypotrichosis simplex. *Hum. Mutat.* **32,** 710–714 (2011).
133 Le Goff, C., Mahaut, C., Wang, L. W., Allali, S., Abhyankar, A., Jensen, S. *et al.* Mutations in the TGFbeta binding-protein-like domain 5 of FBN1 are responsible for acromicric and geleophysic dysplasias. *Am. J. Hum. Genet.* **89,** 7–14 (2011).
134 Le Goff, C., Mahaut, C., Abhyankar, A., Le Goff, W., Serre, V., Afenjar, A. *et al.* Mutations at a single codon in Mad homology 2 domain of SMAD4 cause Myhre syndrome. *Nat. Genet.* **44,** 85–88 (2011).
135 Tsurusaki, Y., Osaka, H., Hamanoue, H., Shimbo, H., Tsuji, M., Doi, H. *et al.* Rapid detection of a mutation causing X-linked leucoencephalopathy by exome sequencing. *J. Med. Genet.* **48,** 606–609 (2011).
136 Shamseldin, H. E., Faden, M. A., Alashram, W. & Alkuraya, F. S. Identification of a novel DLX5 mutation in a family with autosomal recessive split hand and foot malformation. *J. Med. Genet.* **49,** 16–20 (2012).
137 Khan, K., Logan, C. V., McKibbin, M., Sheridan, E., Elcioglu, N. H., Yenice, O. *et al.* Next generation sequencing identifies mutations in Atonal homolog 7 (ATOH7) in families with global eye developmental defects. *Hum. Mol. Genet.* **21,** 776–783 (2012).
138 Zhang, Z., Xia, W., He, J., Ke, Y., Yue, H., Wang, C. *et al.* Exome sequencing identifies SLCO2A1 mutations as a cause of primary hypertrophic osteoarthropathy. *Am. J. Hum. Genet.* **90,** 125–132 (2012).
139 Mitchell, K., O'Sullivan, J., Missero, C., Blair, E., Richardson, R., Anderson, B. *et al.* Exome sequence identifies RIPK4 as the bartsocas—Papas syndrome locus. *Am. J. Hum. Genet.* **90,** 69–75 (2012).
140 Walne, A., Dokal, A., Plagnol, V., Beswick, R., Kirwan, M., de la Fuente, J. *et al.* Exome sequencing identifies MPL as a causative gene in familial aplastic anemia. *Haematologica* **97,** 524–528 (2012).
141 Cabral, R. M., Kurban, M., Wajid, M., Shimomura, Y., Petukhova, L. & Christiano, A. M. Whole-exome sequencing in a single proband reveals a mutation in the CHST8 gene in autosomal recessive peeling skin syndrome. *Genomics* **99,** 202–208 (2012).
142 Mayr, J. A., Haack, T. B., Graf, E., Zimmermann, F. A., Wieland, T., Haberberger, B. *et al.* Lack of the mitochondrial protein acylglycerol kinase causes Sengers syndrome. *Am. J. Hum. Genet.* **90,** 314–320 (2012).
143 Boyden, L. M., Choi, M., Choate, K. A., Nelson-Williams, C. J., Farhi, A., Toka, H. R. *et al.* Mutations in kelch-like 3 and cullin 3 cause hypertension and electrolyte abnormalities. *Nature* **482,** 98–102 (2012).
144 Gibson, W. T., Hood, R. L., Zhan, S. H., Bulman, D. E., Fejes, A. P., Moore, R. *et al.* Mutations in EZH2 cause Weaver syndrome. *Am. J. Hum. Genet.* **90,** 110–118 (2012).
145 Simpson, M. A., Deshpande, C., Dafou, D., Vissers, L. E., Woollard, W. J., Holder, S. E. *et al. De Novo* mutations of the gene encoding the histone acetyltransferase KAT6B cause genitopatellar syndrome. *Am. J. Hum. Genet.* **90,** 290–294 (2012).
146 Campeau, P. M., Kim, J. C., Lu, J. T., Schwartzentruber, J. A., Abdul-Rahman, O. A., Schlaubitz, S. *et al.* Mutations in KAT6B, encoding a histone acetyltransferase, cause genitopatellar syndrome. *Am. J. Hum. Genet.* **90,** 282–289 (2012).
147 Bochukova, E., Schoenmakers, N., Agostini, M., Schoenmakers, E., Rajanayagam, O., Keogh, J. M. *et al.* A mutation in the thyroid hormone receptor alpha gene. *N. Engl. J. Med.* **366,** 243–249 (2012).
148 Hood, R. L., Lines, M. A., Nikkel, S. M., Schwartzentruber, J., Beaulieu, C., Nowaczyk, M. J. *et al.* Mutations in SRCAP, Encoding SNF2-Related CREBBP Activator Protein, Cause Floating-Harbor Syndrome. *Am. J. Hum. Genet.* **90,** 308–313 (2012).
149 Montenegro, G., Rebelo, A. P., Connell, J., Allison, R., Babalini, C., D'Aloia, M. *et al.* Mutations in the ER-shaping protein reticulon 2 cause the axon-degenerative disorder hereditary spastic paraplegia type 12. *J. Clin. Invest.* **122,** 538–544 (2012).
150 Ostergaard, P., Simpson, M. A., Mendola, A., Vasudevan, P., Connell, F. C., van Impel, A. *et al.* Mutations in KIF11 cause autosomal-dominant microcephaly variably associated with congenital lymphedema and chorioretinopathy. *Am. J. Hum. Genet.* **90,** 356–362 (2012).
151 Jones, M. A., Ng, B. G., Bhide, S., Chin, E., Rhodenizer, D., He, P. *et al.* DDOST mutations identified by whole-exome sequencing are implicated in congenital disorders of glycosylation. *Am. J. Hum. Genet.* **90,** 363–368 (2012).
152 Lines, M. A., Huang, L., Schwartzentruber, J., Douglas, S. L., Lynch, D. C., Beaulieu, C. *et al.* Haploinsufficiency of a spliceosomal GTPase encoded by EFTUD2 causes mandibulofacial dysostosis with microcephaly. *Am. J. Hum. Genet.* **90,** 369–377 (2012).
153 Harms, M. B., Sommerville, R. B., Allred, P., Bell, S., Ma, D., Cooper, P. *et al.* Exome sequencing reveals DNAJB6 mutations in dominantly-inherited myopathy. *Ann. Neurol.* **71,** 407–416 (2012).
154 Audo, I., Bujakowska, K., Orhan, E., Poloschek, C. M., Defoort-Dhellemmes, S., Drumare, I. *et al.* Whole-exome sequencing identifies mutations in GPR179 leading to autosomal-recessive complete congenital stationary night blindness. *Am. J. Hum. Genet.* **90,** 321–330 (2012).
155 Hussain, M. S., Baig, S. M., Neumann, S., Nurnberg, G., Farooq, M., Ahmad, I. *et al.* A truncating mutation of CEP135 causes primary microcephaly and disturbed centrosomal function. *Am. J. Hum. Genet.* **90,** 871–878 (2012).
156 Kirwan, M., Walne, A. J., Plagnol, V., Velangi, M., Ho, A., Hossain, U. *et al.* Exome sequencing identifies autosomal-dominant SRP72 mutations associated with familial aplasia and myelodysplasia. *Am. J. Hum. Genet.* **90,** 888–892 (2012).
157 Lee, H., Graham, Jr. J. M., Rimoin, D. L., Lachman, R. S., Krejci, P., Tompson, S. W. *et al.* Exome sequencing identifies PDE4D mutations in acrodysostosis. *Am. J. Hum. Genet.* **90,** 746–751 (2012).

158 Lin, Z., Chen, Q., Lee, M., Cao, X., Zhang, J., Ma, D. *et al.* Exome sequencing reveals mutations in TRPV3 as a cause of Olmsted syndrome. *Am. J. Hum. Genet.* **90,** 558–564 (2012).
159 Fiskerstrand, T., Arshad, N., Haukanes, B. I., Tronstad, R. R., Pham, K. D., Johansson, S. *et al.* Familial diarrhea syndrome caused by an activating GUCY2C mutation. *N. Engl. J. Med.* **366,** 1586–1595 (2012).
160 Bernier, F. P., Caluseriu, O., Ng, S., Schwartzentruber, J., Buckingham, K. J., Innes, A. M. *et al.* Haploinsufficiency of SF3B4, a component of the pre-mrna spliceosomal complex, causes Nager syndrome. *Am. J. Hum. Genet.* **90,** 925–933 (2012).
161 Spiegel, R., Pines, O., Ta-Shma, A., Burak, E., Shaag, A., Halvardson, J. *et al.* Infantile cerebellar-retinal degeneration associated with a mutation in mitochondrial aconitase, ACO2. *Am. J. Hum. Genet.* **90,** 518–523 (2012).
162 Zankl, A., Duncan, E. L., Leo, P. J., Clark, G. R., Glazov, E. A., Addor, M. C. *et al.* Multicentric carpotarsal osteolysis is caused by mutations clustering in the amino-terminal transcriptional activation domain of MAFB. *Am. J. Hum. Genet.* **90,** 494–501 (2012).
163 Tsurusaki, Y., Okamoto, N., Ohashi, H., Kosho, T., Imai, Y., Hibi-Ko, Y. *et al.* Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. *Nat. Genet.* **44,** 376–378 (2012).
164 Srour, M., Schwartzentruber, J., Hamdan, F. F., Ospina, L. H., Patry, L., Labuda, D. *et al.* Mutations in C5ORF42 cause Joubert syndrome in the French Canadian population. *Am. J. Hum. Genet.* **90,** 693–700 (2012).
165 Polvi, A., Linnankivi, T., Kivela, T., Herva, R., Keating, J. P., Makitie, O. *et al.* Mutations in CTC1, encoding the CTS telomere maintenance complex component 1, cause cerebroretinal microangiopathy with calcifications and cysts. *Am. J. Hum. Genet.* **90,** 540–549 (2012).
166 Schossig, A., Wolf, N. I., Fischer, C., Fischer, M., Stocker, G., Pabinger, S. *et al.* Mutations in ROGDI cause Kohlschutter-Tonz syndrome. *Am. J. Hum. Genet.* **90,** 701–707 (2012).
167 Santen, G. W., Aten, E., Sun, Y., Almomani, R., Gilissen, C., Nielsen, M. *et al.* Mutations in SWI/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome. *Nat. Genet.* **44,** 379–380 (2012).
168 Nakazawa, Y., Sasaki, K., Mitsutake, N., Matsuse, M., Shimada, M., Nardo, T. *et al.* Mutations in UVSSA cause UV-sensitive syndrome and impair RNA polymerase IIo processing in transcription-coupled nucleotide-excision repair. *Nat. Genet.* **44,** 586–592 (2012).
169 Austin, E. D., Ma, L., Leduc, C., Berman Rosenzweig, E., Borczuk, A., Phillips, J. A. *et al.* Whole exome sequencing to identify a novel gene (caveolin-1) associated with human pulmonary arterial hypertension. *Circ. Cardiovasc. Genet.* **44,** 586–592 (2012).