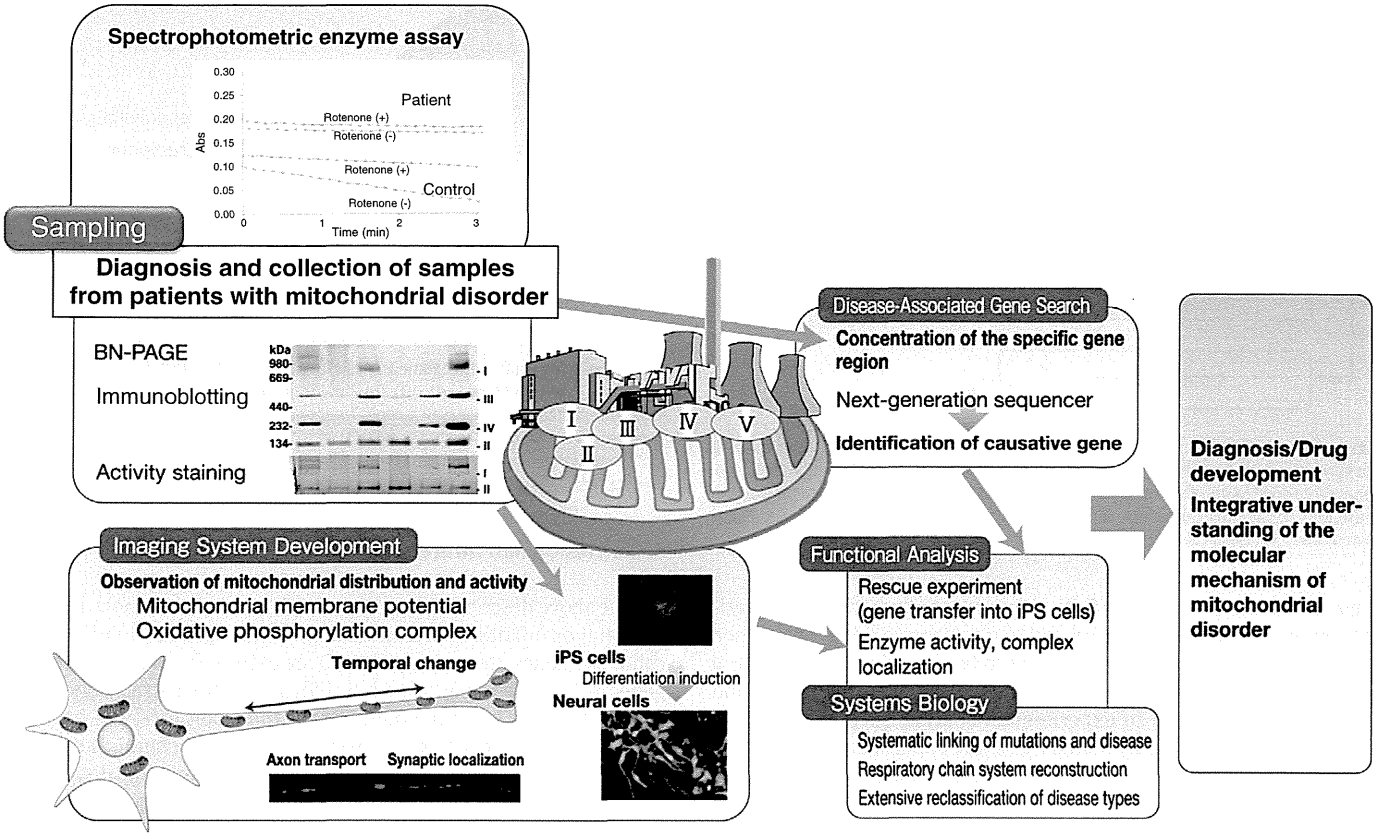

Fig. 1. Outline of exome analysis project for MRCD patients. The first step is 'Sampling', which refers to diagnosis and collection of samples from patients with mitochondrial disorders using both spectrophotometric enzyme assay [11] and BN-PAGE [13]. The next step is 'Disease-Associated Gene Search' using exome analysis. In 'Functional Analysis' and 'System Biology', candidate causal genes are confirmed at the cellular level by rescue experiment or other means. In 'Imaging System Development', induced pluripotent stem cells are created using fibroblasts and differentiated into neurons and glia cells to reproduce the pathology of mitochondrial dysfunction. The final purpose of our project is integrative understanding of the molecular mechanisms of mitochondrial disorders.

inheritance (as shown in Table 1), we prioritized such genes as harboring rare variants in a homozygous or compound heterozygous fashion. Low priority is given to the analysis of genes showing mutation in only one allele because patients and healthy control individuals harbored a comparable number of rare heterozygous alleles; we were unable to prioritize dominant-acting genes.

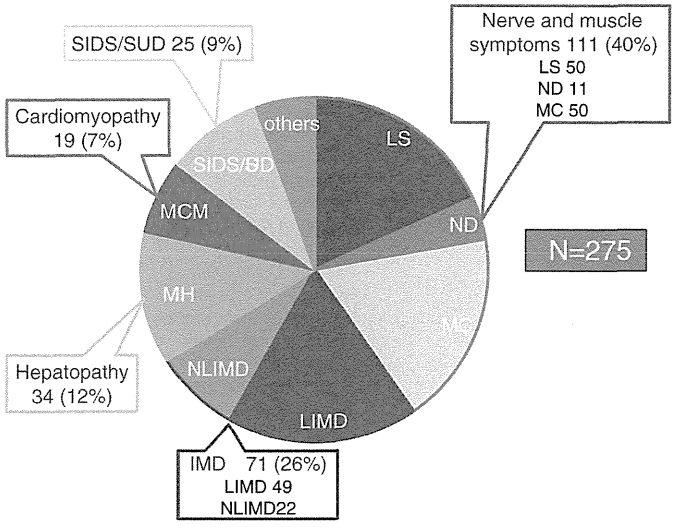

Fig. 2. Breakdown of clinical diagnoses of mitochondrial disorders in our institute as of January 2013. LS, Leigh syndrome; ND, neurodegenerative disorder; MC, mitochondrial cytopathy; IMD, infantile mitochondrial disease (lethal and non-lethal); MH, mitochondrial hepatopathy; MCM, mitochondrial cardiomyopathy; SIDS, sudden infant death syndrome; SUD, sudden unexpected death.

Our current bioinformatics analysis pipeline is as follows: read alignment was performed with a Burrows–Wheeler Aligner (BWA, version 0.7.0) [17] using the 1000 Genomes project phase II reference genome (hs37d5.fa). PCR duplicate reads were removed using Picard (version 1.89) (http://picard.sourceforge.net) and non-mappable reads were removed using SAMtools (version 0.1.19) [18]. After filtering out these reads, the Genome Analysis Toolkit (GATK) version 2.4-9-nightly-2013-04-12-g3fc5478 [19] was used to realign insertions and deletions, and for quality recalibration and variant calling (UnifiedGenotyper). Detected variants were annotated using ANNOVAR (version 2013Feb21) [20] and custom ruby scripts. The effect of the mutations on protein function was assessed by SIFT and GERP using dbNSFP [21]. The positions of mutations were based on RefSeq transcript sequences. Variants were assessed by comparing allele frequencies in the dbSNP135, Exome Sequencing Project (ESP5400) data set, and 1000 Genomes Projects (based on phase 1 release v3 called from 20101123 alignments). As mitochondrial disorders are rare, we excluded variants present in dbSNP with a frequency > 0.1%. After filtering out these variants, the VAAST program [22] was used to create a candidate gene list in each patient showing recessive characteristics.

As stated above, because mitochondrial disease patients have very high heterogeneity, the number of patients sharing the same gene mutation is quite low. Hence, attention should be directed towards removing these mutations from the disease candidates when the same amino acid substitutions are shared among multiple patients in our study, because these variants are highly likely to be SNPs unique to the Japanese population. Using these criteria, we are able to narrow down the number of variants to a mean of several genes for each patient. After listing

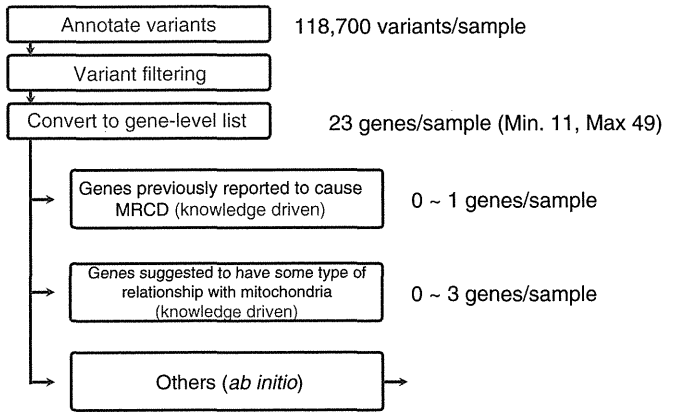

Fig. 3. Narrowing down of gene mutations discovered by exome analysis. After filtering out variants with the methods described in the 'Exome analysis of MRCD patients' section, genes were divided into three categories: (1) those that have previously been reported to cause MRCD; (2) those for which some relationship with mitochondria has been suggested; and (3) others (*ab initio*).

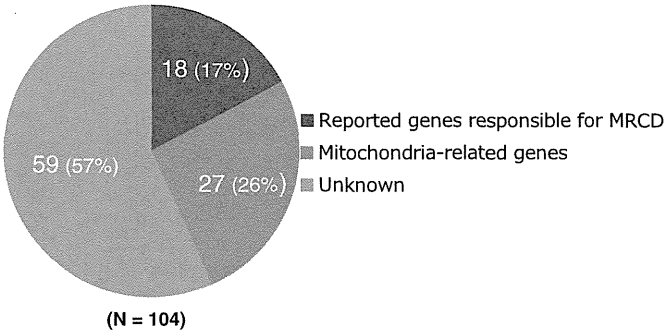

Fig. 4. Candidate genes with exome analysis for MRCD patients. Results of analysis for 104 patients to date (as of January 2013) are shown. Eighteen patients (17%) had variants previously reported to be disease-causing. Twenty-seven patients (26%) had mutations in genes suggested to be associated somehow with mitochondria. The remaining 59 patients (57%) are highly likely to contain completely novel disease-causing mutations for which no data have been obtained to date.

these candidate variants, we further investigated whether these variants are located within genes related to mitochondrial function. When genes overlapped with those reported to be related to mitochondrial function, we found that they were likely to be causative genes and were further subjected to experimental analysis such as haplotype phasing or functional assay including rescue experiments. To prepare a list of genes reported to be related to mitochondria, we included genes annotated as somehow related to mitochondria in the UniProt (http://www.uniprot.org/) [23] database, as well as the MitoCarta database (http://www.broadinstitute.org/pubs/MitoCarta/index.html) [24], which includes approximately 1000 gene products listed with the use of shotgun proteomics and mitochondrial localization analysis.

We also investigated whether there is Long Contiguous Stretch of Homozygosity (LCSH) using Affymetrix SNP arrays in a majority of patients. Although no cases of consanguineous marriage were reported in the interviews with the primary physician, about 5% of cases harbor LCSH proven by SNP arrays. When homozygous mutations are localized in these LCSH regions, the mutations are highly likely to be causative of disease.

5. Results of exome analysis for MRCD patients

The variants (mutations) found in the process of narrowing down the gene mutations discovered to date are shown in Fig. 3. These genes were narrowed down to the final candidate genes and divided into three categories: (1) those that have previously been reported to cause MRCD; (2) those for which some relationship with mitochondria has been suggested; and (3) others (*ab initio*). The results of analysis of 104 patients to date (as of January 2013) are shown in Fig. 4. Eighteen patients (17%) had variants previously reported to be disease-causing. Among these 18 patients, one had a homozygote of a previously reported mutation and two had a compound heterozygote of a reported and a novel mutation (data not shown). All other mutations found in this study were new. Twenty-seven patients (26%) had mutations in genes suggested to be associated somehow with mitochondria, and it is likely that they are novel disease-causing genes in mitochondrial disorders. Table 2 lists the functions of the genes in these 27 cases. For the remaining 59 cases, each patient has about 20 gene variants that are unique to each patient, and it is necessary to confirm whether any of these mutations can actually cause the disease. These 59 patients are highly likely to contain completely novel disease-causing mutations for which no clues have been obtained to date. The biggest issue we currently face is how to confirm the disease-causing gene from these 20 gene variants for each patient.

6. Conclusion and future prospects

The above describes the progress we have made in exome analysis of neonatal or infantile MRCD patients. While we have identified many candidate genes, the causes of MRCD are extremely diverse and heterogeneous. Thus, in many cases, it is difficult to demonstrate conclusively that a mutation in a candidate gene is the true cause. We have performed analyses focusing on cases in which a biochemical diagnosis was established at the cellular level in addition to clinical symptoms such as enzyme activity and complex formation abnormalities. Nonetheless, confirmation of the causal genes with rescue experiments or other means is difficult. In the future, it will be necessary to increase the case number or search for patients with similar symptoms and similar gene mutations in collaboration with researchers throughout the world. We are currently conducting analyses of pediatric patients with a focus on MRCD, and gene mutations (amino acid substitutions) harbored by patients of the childhood onset type are probably variants conferring major damage on enzyme activity or protein function. Onset is also thought to occur in adulthood rather than in childhood in some cases of milder (hypomorphic: partial loss of function) variants with the same gene defect. As these are thought to include nerve diseases,

Table 2
Functions of new disease-causing candidate genes for MRCD.

MtoX#1	Non-receptor tyrosine kinase
MtoX#2	Acyl-CoA thioesterase
MtoX#3	Fatty acid β oxidation
MtoX#4	tRNA synthetase
MtoX#5	ABC transporter superfamily
MtoX#6	ATR-dependent AMP-binding enzyme family
MtoX#7	Heme biosynthesis
MtoX#8	AAA ATPase family
MtoX#9	Pre-mRNA splicing factor
MtoX#10	Creatine kinase
MtoX#11	Synaptic transmission
MtoX#12	Synthesis of Coenzyme Q
MtoX#13	Heme biosynthetic process
MtoX#14	Citrate synthase family.
MtoX#15	Cholesterol metabolism
MtoX#16	Mitochondrial fission
MtoX#17	Muscle organ development
MtoX#18	Cholesterol biosynthetic process
MtoX#19	Ribosomal protein
MtoX#20	Tumor suppressor
MtoX#21	A component of complex I
MtoX#22	A protease, located in inner membrane
MtoX#23	Regulation of PDH
MtoX#24	Mitochondrial translation
MtoX#25	Queuosine biosynthetic process
MtoX#26	Mitochondrial carrier family
MtoX#27	Methyltransferase superfamilya

mental disorders, and diabetes or other metabolic diseases of unknown cause, we plan to conduct research based on the assumption that such cases include those caused by abnormalities in genes identified in MRCD patients.

Acknowledgements

This work was supported in part by a grant for Innovative Cell Biology by Innovative Technology (Cell Innovation Program) and Support Project from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan, by a grant for Strategic Research Centers in Private Universities from MEXT, Japan to Saitama Medical University Research Center for Genomic Medicine, and by Grants-in-Aid for Research on Intractable Diseases (Mitochondrial Disorder) from the Ministry of Health, Labor and Welfare of Japan. Dr. Murayama was supported by the Kawano Masanori Memorial Public Interest Incorporated Foundation for Promotion of Pediatrics. The authors would also like to thank Dr. Ayako Fujinami, Dr. Kaori Muta, Dr. Emi Kawachi, Dr. Takuya Fushimi, Dr. Keiko Ichimoto, Dr. Tomoko Tsuruoka, Ms. Keio Baba and Ms. Masami Ajima at Chiba Children's Hospital for their support.

References

[1] D. Skladal, J. Halliday, D.R. Thorburn, Minimum birth prevalence of mitochondrial respiratory chain disorders in children, Brain 126 (2003) 1905–1912.
[2] S. Dimauro, G. Davidzon, Mitochondrial DNA and disease, Ann. Med. 37 (2005) 222–232.
[3] D.M. Kirby, D.R. Thorburn, Approaches to finding the molecular basis of mitochondrial oxidative phosphorylation disorders, Twin Res. Hum. Genet. 11 (2008) 395–411.
[4] D.C. Wallace, Mitochondria and cancer, Nat. Rev. Cancer 12 (2012) 685–698.
[5] M.F. Lopez, B.S. Kristal, E. Chernokalskaya, A. Lazarev, A.I. Shestopalov, A. Bogdanova, M. Robinson, High-throughput profiling of the mitochondrial proteome using affinity fractionation and automation, Electrophoresis 21 (2000) 3427–3440.
[6] H. Schägger, K. Pfeiffer, The ratio of oxidative phosphorylation complexes I–V in bovine heart mitochondria and the composition of respiratory chain supercomplexes, J. Biol. Chem. 276 (2001) 37861–37867.
[7] D.R. Thorburn, Mitochondrial disorders: prevalence, myths and advances, J. Inherit. Metab. Dis. 27 (2004) 349–362.
[8] W.J. Koopman, P.H. Willems, J.A. Smeitink, Monogenic mitochondrial disorders, N. Engl. J. Med. 366 (2012) 1132–1141.
[9] S.B. Vafai, V.K. Mootha, Mitochondrial disorders as windows into an ancient organelle, Nature 491 (2012) 374–383.
[10] E.J. Tucker, A.G. Compton, D.R. Thorburn, Recent advances in the genetics of mitochondrial encephalopathies, Curr. Neurol. Neurosci. Rep. 10 (2010) 277–285.
[11] D.M. Kirby, M. Crawford, M.A. Cleary, H.H. Dahl, X. Dennett, D.R. Thorburn, Respiratory chain complex I deficiency: an underdiagnosed energy generation disorder, Neurology 52 (1999) 1255–1264.
[12] F.P. Bernier, A. Boneh, X. Dennett, C.W. Chow, M.A. Cleary, D.R. Thorburn, Diagnostic criteria for respiratory chain disorders in adults and children, Neurology 59 (2002) 1406–1411.
[13] H. Schägger, G. von Jagow, Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form, Anal. Biochem. 199 (1991) 223–231.
[14] K. Gibson, J.L. Halliday, D.M. Kirby, J. Yaplito-Lee, D.R. Thorburn, A. Boneh, Mitochondrial oxidative phosphorylation disorders presenting in neonates: clinical manifestations and enzymatic and molecular diagnoses, Pediatrics 122 (2008) 1003–1008.
[15] T. Yamazaki, K. Murayama, A.G. Compton, C. Sugiana, H. Harashima, S. Amemiya, M. Ajima, T. Tsuruoka, A. Fujinami, E. Kawachi, Y. Kurashige, K. Matsushita, H. Wakiguchi, M. Mori, H. Iwasa, Y. Okazaki, D.R. Thorburn, A. Ohtake, Molecular diagnosis of mitochondrial respiratory chain disorders in Japan: focusing on mitochondrial DNA depletion syndrome, Pediatr. Int. 56 (2014)(in press).
[16] C. Gilissen, A. Hoischen, H.G. Brunner, J.A. Veltman, Disease gene identification strategies for exome sequencing, Eur. J. Hum. Genet. 20 (2012) 490–497.
[17] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform, Bioinformatics 25 (2009) 1754–1760.
[18] S.B. Ng, E.H. Turner, P.D. Robertson, S.D. Flygare, A.W. Bigham, C. Lee, T. Shaffer, M. Wong, A. Bhattacharjee, E.E. Eichler, M. Bamshad, D.A. Nickerson, J. Shendure, Targeted capture and massively parallel sequencing of 12 human exomes, Nature 461 (2009) 272–276.
[19] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M.A. DePristo, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, Genome Res. 20 (2010) 1297–1303.
[20] K. Wang, M. Li, H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data, Nucleic Acids Res. 38 (2010) e164.
[21] X. Liu, X. Jian, E. Boerwinkle, dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions, Hum. Mutat. 32 (2011) 894–899.
[22] M. Yandell, C. Huff, H. Hu, M. Singleton, B. Moore, J. Xing, L.B. Jorde, M.G. Reese, A probabilistic disease-gene finder for personal genomes, Genome Res. 21 (2011) 1529–1542.
[23] The UniProt Consortium, Update on activities at the Universal Protein Resource (UniProt) in 2013, Nucleic Acids Res. 41 (2013) D43–D47.
[24] D.J. Pagliarini, S.E. Calvo, B. Chang, S.A. Sheth, S.B. Vafai, S.E. Ong, G.A. Walford, C. Sugiana, A. Boneh, W.K. Chen, D.E. Hill, M. Vidal, J.G. Evans, D.R. Thorburn, S.A. Carr, V.K. Mootha, A mitochondrial protein compendium elucidates complex I disease biology, Cell 134 (2008) 112–123.