Figure 2 | The impact of very-rare variants on genomic regions and functional categories. (a) The SFSs of intergenic region for SNVs of 1KJPN (blue) and 1KGP (red). (b) The numbers of SNVs observed in 1KJPN and 1KGP are depicted as four functional categories. The fraction of very-rare variants observed in 1KJPN are depicted with 95% binomial confidence interval according to (c) genomic region, (d) probable consequences for coding regions, (e) in noncoding regions and (f) for scaled C scores. Because the number of genotyped individuals in the confidence SNVs is different among sites because of the individual depth filter, we applied a hypergeometric projection[65], which subsamples each variant down to a sample size of 963 (90% of 1,070 samples) to obtain the SFSs of the confidence SNVs for a,c–f.

illustrate that this reported sequencing paradigm successfully discovered biologically relevant SNVs, most of which are rare in the Japanese population.

Structural variants. High-coverage sequencing data allowed us to catalogue both novel and known deletions and insertions within 1KJPN (Fig. 3a). The size-frequency spectrum showed that larger indels were less abundant than smaller ones, consistent with observations from both 1KGP[9] and Genome of the Netherlands[20]. Notably, most of the longer insertions (> 10 bp) detected in this pipeline[35] were novel, which represents the usefulness of this high-coverage sequencing strategy (see Methods).
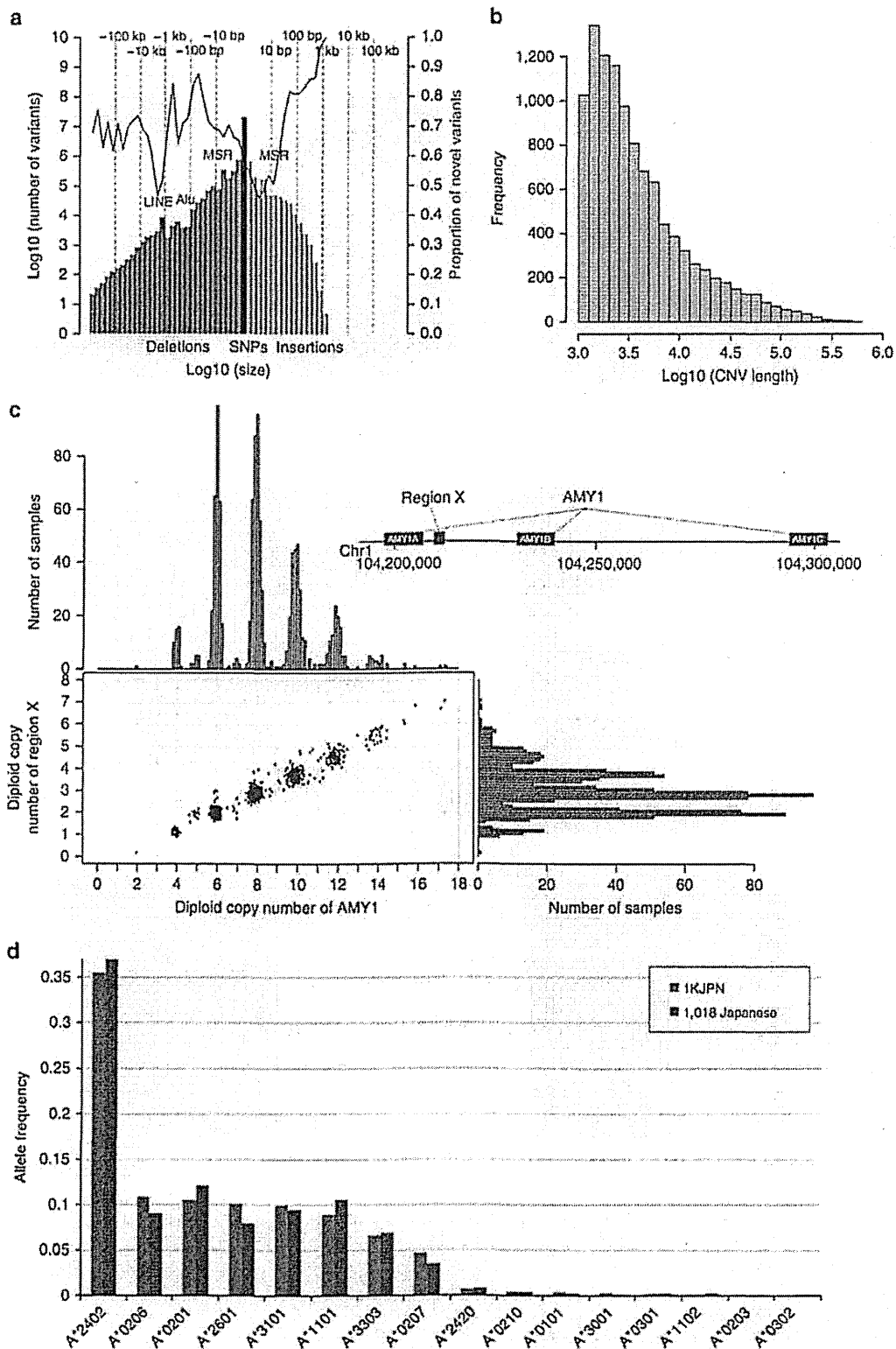
— 355 —

Figure 3 | Properties of genomic variation discovered in 1KJPN. (a) The size-frequency spectrum of SNVs, deletions and insertions discovered by high-coverage sequencing in 1KJPN. Novelty rates are shown by the red line. Peaks corresponding to long interspersed elements (LINE), Alu and microsatellite repeat (MSR) are shown. (b) Size-frequency spectrum of CNVs estimated from high-coverage sequencing data in the genic regions in 1KJPN. (c) Histograms and scatterplot of diploid copy numbers of AMY1 genes (blue) and region X (red) in 1KJPN. A diagram depicting the positions of AMY1A, Region X, AMY1B and AMY1C on chromosome 1 of GRCh37 is shown in the right top. (d) Allele frequencies for HLA-A in 1,070 individuals in 1KJPN estimated by high-coverage sequencing (blue), and 1,018 Japanese individuals typed by PCR-SSOP (red)[42].

— 356 —

We also constructed a genome-wide map of CNVs overlapping with the genic region in 1KJPN, in which 25,923 CNV loci were identified; precise copy numbers were quantified on the basis of the alignment status and depth of coverage of sequence reads on individual CNV loci (Fig. 3b, Methods and Supplementary Data 1).

Among the revealed CNVs, we found that most Japanese individuals have more than two diploid copies of the salivary amylase gene (AMY1; Fig. 3c). The mean diploid copy number of AMY1 within 1KJPN was 8.27, which is significantly higher than the number reported in populations who consume low amounts of starch (mean 5.44, $N = 93$)[18]. This observation is consistent with the higher copy number of AMY1 reported in high-starch populations, suggesting adaptation of an increased copy number according to dietary habits[18]. Interestingly, we found that the copy-number unit of AMY1 was two (that is, even AMY1 diploid copy numbers), in which the closely linked AMY1A and AMY1B loci duplicated together (Fig. 3c). Indeed, the diploid copy number of AMY1 is proportional to the copy number of intermediate region between AMY1A and AMY1B (designated by 'Region X' in the schematic diagram of the human reference genome in Fig. 3c), in which the relationship can be described as $y = 2n + 2$, as the diplotype model, where y is the diploid copy number of AMY1, and n is the diploid copy number of Region X. We also performed the digital PCR analysis to validate the estimated copy numbers in 10 samples (Methods and Supplementary Fig. 5). This means that two AMY1 gene loci, AMY1A and AMY1B, and Region X are close to one another on a chromosome, and this region as a whole is the copy unit of the AMY1 gene. We further confirmed that there is almost no variation in copy number outside of the AMY1 genes. For example, diploid copy numbers of AMY2B and Region Y are mostly two (Supplementary Fig. 6b). This observation is consistent with the previously proposed haplotype model for the salivary amylase gene, $(AMY1A\text{-}AMY1B)n\text{-}AMY1C$[36]. Since the copy number of CNV loci present in a sample (gene dosage) is likely to affect the relative abundance of gene expression, this information will be useful for finding association of genetic variants with phenotypes or diseases in future follow-up studies. The CNVs reported here can be used to investigate associations with human disease phenotypes as previously reported[37,38].

HLA types.

Another important aspect of deep sequencing is the ability to determine highly polymorphic gene alleles. Human leucocyte antigen (HLA) genes are highly polymorphic in the human genome, and their haplotype structures are different among human populations[39]. In this study, we have typed HLA alleles of HLA-A, HLA-B and HLA-C loci of 1KJPN using newly developed software called HLA-VBSeq[40]. The method is based on the alignment of sequence reads to the genomic HLA allele sequences registered in the IMGT/HLA database[41]. We successfully typed HLA-A alleles for most of the individuals in 1KJPN (2,063 out of 2,140 alleles) at full resolution (8-digit in HLA nomenclature). We observed that the frequencies of HLA-A, -B and -C alleles at 4-digit resolution (nucleotide differences that change amino-acid sequences) estimated in our study were very similar to previously published frequencies (Fig. 3d and Supplementary Fig. 6c,d), in which HLA types were determined using PCR-SSOP among a different set of 1,018 Japanese individuals[42]. Because HLA genes are crucial in determining the outcome of organ transplantation and susceptibility to infectious and autoimmune diseases[43,44], this HLA typing will be particularly important for GWAS[45] and clinical practice, such as donor–recipient matching[46].

Haplotyping and imputation.

Genome-wide genotype imputation is a statistical technique to estimate untyped genotypes from known haplotype information, which is cost-effective for GWAS with SNP arrays when compared with exome sequencing and WGS. We constructed a phased reference panel with high-sensitive SNVs plus short insertions and deletions in 1KJPN using HapMonster[47] and ShapeIT2 (ref. 48). The performance of genotype imputation with the reference panel for 131 Japanese individuals (who were not among the individuals used to compile 1KJPN and not cryptic relatives to them; estimated identity-by-descent <0.125) was assessed by comparing their imputed genotypes and those obtained with the same sequencing protocol and variant calling pipeline that was used to generate 1KJPN. For the imputation, genotypes at sites designed in Omni2.5 were used (Supplementary Methods). We compared 1KJPN and the following three types of reference panels for genotype imputation: (i) the 1,092 cosmopolitan samples in 1KGP, (ii) the 89 JPT samples in 1KGP and (iii) the 1KJPN plus 1KGP. From the comparison, the highest r^2 value (the measure of imputation accuracy) was achieved with 1KJPN plus 1KGP, especially in variants with MAF ≤ 5%; the mean r^2 values were 0.47 for very-rare SNVs, 0.66 for rare SNVs and 0.78 for low frequency SNVs (Fig. 4a). The significant improvement of imputation accuracy using the 1KJPN data suggests the importance of construction and examination of a population-specific reference panel.

To illustrate the effectiveness of the genotype imputation strategy in GWAS, we performed imputation from genotyped data set in a case–control study of MMD[49] with 1KJPN (Supplementary Methods). MMD is a progressive cerebrovascular disorder caused by blocked arteries at the base of the brain and has unusually high prevalence among the Japanese. The data set contains the genotypes with Illumina HumanOmni1-Quad BeadChip for 72 Japanese MMD patients and 45 healthy controls from HapMap JPT collection. We performed GWAS for the imputed data sets with the same statistical parameters used in the original study[49]. A synonymous SNP rs11870849 located at the coding region of ENDOV (chr17:78,411,073) with a P value of 6.95×10^{-9} (χ^2-test) was identified as the highest association SNP (Fig. 4c top) from the original (non-imputed) data set (Fig. 4b). In contrast, by the imputation employing 1KJPN (Fig. 4c bottom), a nonsynonymous SNP rs112735431 located in RNF213 (chr17:78,358,945; MAF = 0.0089 in 1KJPN) was identified as the highest association SNP with a P value of 8.07×10^{-10} over a significance threshold of a P value $< 5.06 \times 10^{-9}$. This SNP was confirmed by fine mapping of chromosome 17q ter locus with another set of SNPs followed by targeted sequencing[49], which was suggested to be a major MMD causal variant[49,50]. Thus, we have directly inferred the causal SNP with our 1KJPN reference panel.

Functional variant load in individuals.

We summarized the number of disease-causing variants in autosomes per individual for each derived allele frequency (Table 2 and Supplementary Methods). On average, one individual has 11.2 disease-causing alleles in terms of disease-causing mutation (DM) category of HGMD[31] (9.6 as heterozygous and 1.6 as homozygous in the high-confidence SNVs). Similarly, we calculated the number of nonsense (stop-gained) variants per individual on the high-confidence SNVs. Each individual has on average 41.6 heterozygous and 12.1 homozygous stop-gained SNVs. These estimates are reasonably consistent with those in East Asian populations in 1KGP (JPT, CHB and CHS). Although there are reports on higher mutational loads in founding populations[13,21,51], our analysis of mutational load did not
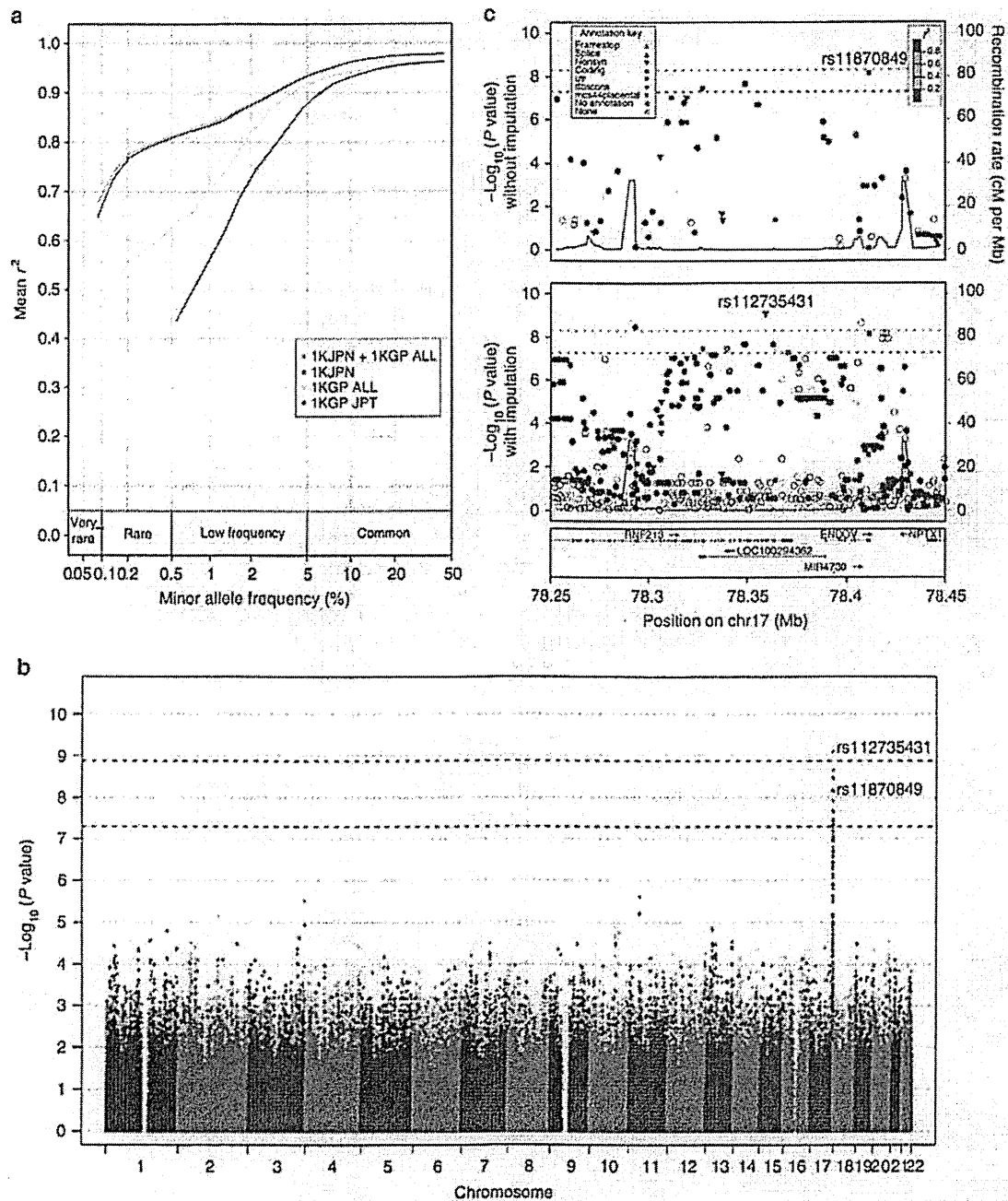
Figure 4 | Imputation with the Japanese reference panel. (a) Comparison of imputation performance (r^2) for four reference panels: 1,070 individuals in 1KJPN (1KJPN), 1,092 cosmopolitan samples in 1KGP (1KGP ALL), 1KJPN plus 1KGP ALL (1KJPN + 1KGP ALL) and 89 Japanese individuals in 1KGP (1KGP JPT). The x axis represents the MAF of each panel. The y axis represents the averaged r^2 at SNV sites that exist in both the cosmopolitan samples of 1KGP and 1KJPN. (b) A Manhattan plot of P values from GWAS of MMD. The SNV sites from the original data set and imputed markers are plotted as dots in magenta and grey, respectively. Blue and red lines display the significance threshold of the original and imputed results, respectively. Only one significant signal was identified on chromosome 17. (c) A plot of P values from GWAS of MMD with the original (non-imputed; upper panel) and imputed (lower panel) data set around the SNP exhibiting the significant signal in b. In the imputed result, the SNP with the highest association is a nonsynonymous variant of RNF213, and was reported as one of the MMD-causing variants in the original study. In contrast, from the non-imputed result the SNP with the highest association is located in the coding region of ENDOV.

show clear difference between Japanese and other East Asian populations. In terms of nonsense mutation, the estimate of individual load in 1KJPN is consistent with those in the analysis of the Dutch population[20]. In 1KJPN, 3,505 SNVs were annotated as stop-gained, while among them only 4.5% were annotated as phenotypic variants in HGMD. This low proportion suggests that

the biological or pathological effects of most stop-gained SNVs are yet to be revealed.

Incidence rates for inherited diseases vary among populations[52]. For example, the incidence of inherited metabolic diseases was found to be 1 in 9,330 in Japan, while among those of European ancestry, metabolic diseases were more

Table 2 | Individual variant load in coding regions.

Variant type	Population	Very-rare (<0.1%)		Rare (0.1-0.5%)		Low (0.5-5%)		Common (>5%)		Total	
		Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
Heterozygous											
HGMD-DM*	1KJPN (1,070) high confidence	0.640	0.814	1.039	1.016	4.757	2.246	3.179	1.604	9.619	3.032
	1KJPN (1,070) high-sensitive	0.675	0.848	1.107	1.051	4.905	2.261	4.388	1.787	11.074	3.136
	1KGP JPT (89)	NA	NA	NA	NA	6.270	2.503	4.169	1.829	10.438	2.969
	1KGP CHB (97)	NA	NA	NA	NA	5.536	2.381	4.464	1.921	10.000	3.218
	1KGP CHS (100)	NA	NA	1.470	1.359	4.320	2.049	3.680	1.803	9.470	2.798
Stop-gained[†]	1KJPN (1,070) high-confidence	2.385	1.550	1.563	1.294	8.679	2.486	29.017	4.327	41.644	5.358
	1KJPN (1,070) high-sensitive	2.624	1.616	1.777	1.376	6.008	2.402	42.125	4.878	52.535	5.795
	1KGP JPT (89)	NA	NA	NA	NA	8.685	2.987	39.337	5.261	48.022	6.166
	1KGP CHB (97)	NA	NA	NA	NA	9.742	3.215	37.845	5.593	47.588	6.777
	1KGP CHS (100)	NA	NA	3.860	2.433	6.580	3.085	36.070	4.860	46.510	5.947
Homozygous											
HGMD-DM*	1KJPN (1070) high-confidence	0.001	0.031	0.003	0.053	0.048	0.230	1.570	1.126	1.621	1.145
	1KJPN (1,070) high-sensitive	0.000	0.000	0.003	0.053	0.050	0.234	1.862	1.235	1.914	1.251
	1KGP JPT (89)	NA	NA	NA	NA	0.022	0.149	1.899	1.244	1.921	1.227
	1KGP CHB (97)	NA	NA	NA	NA	0.052	0.222	2.021	0.989	2.072	1.003
	1KGP CHS (100)	NA	NA	0.000	0.000	0.000	0.000	0.000	0.000	2.110	1.118
Stop-gained[†]	1KJPN (1,070) high-confidence	0.005	0.081	0.004	0.061	0.753	0.747	11.303	2.713	12.064	2.813
	1KJPN (1,070) high-sensitive	0.008	0.101	0.008	0.101	0.099	0.302	12.101	2.853	12.217	2.851
	1KGP JPT (89)	NA	NA	NA	NA	0.067	0.252	11.315	2.898	11.382	2.914
	1KGP CHB (97)	NA	NA	NA	NA	0.052	0.222	12.093	2.758	12.144	2.769
	1KGP CHS (100)	NA	NA	0.000	0.000	0.070	0.326	12.900	3.047	12.970	3.096

CHB, Han Chinese in Beijing, China; CHS, Han Chinese South, China; HGMD, Human Gene Mutation Database; JPT, Japanese in Tokyo, Japan; 1KGP, 1000 Genomes Project; 1KJPN, reference panel of 1,070 Japanese individual; NA, not available; ORF, open reading frame; SNV, single-nucleotide variant.
SNV sites with reliable ancestral states were used.
*HGMD-DM (disease-causing) alleles were analysed if they are derived alleles and alternative (non-reference) alleles.
†We selected stop-gained alleles if they are derived alleles and alternative (non-reference) alleles. We discarded stop-gained SNVs if the proportion of truncated ORF is less than 5%.

prevalent (1 in 4,000–5,000)[53]. Information of risk allele frequencies among populations will be useful for estimating penetrance and identifying additional causes of disease, such as other pathogenic loci or environmental factors. Indeed, we found significant differences of allele frequency for HGMD variants between 1KJPN and each of 14 populations in 1KGP. The 2,638 SNVs in HGMD were significantly different (P value $< 10^{-5}$; Fisher's exact test) in at least one population of the 1KGP. Among them, 36 SNVs were annotated as disease-causing mutations. For example, SNP rs1047781 at chromosome 19:49,206,631 (causing an Ile129Phe change) in the *FUT2* gene (encoding fucosyltransferase 2) showed a drastic difference in the non-reference allele frequency between the 1KJPN (0.38) and 1KGP CEU (0.00). The *FUT2* gene encodes an enzyme for secreting ABH antigens, and the locus is known as a classic human secretor locus[54]. The frequency of homozygotes of this variant was 0.141 in 1KJPN, and this is consistent with the fact that ~15% of Japanese individuals are *FUT2* nonsecretors[55]. Recent studies showed that this SNP rs1047781 was associated with the levels of tumour biomarker and vitamin B12 (refs 56,57).

Discussion

From the high-coverage WGS data collected in the ToMMo cohort project, we have constructed 1KJPN, a reference panel of 1,070 Japanese individuals. This panel includes a highly accurate and comprehensive collection of 21.5 million SNVs ranging from very-rare (observed MAF ≤ 0.1%) to common (>5%) variants. Considering the demographic history of 1KJPN inferred from the SFS of intergenic region, the data cover almost all SNVs (99.0%) that exist in the Japanese population having a minor frequency of

0.1% or greater (Fig. 1d). Furthermore, we would find rarer variants if additional samples will be sequenced. For instance, while we have detected 60.5 % of SNVs with MAF > 0.01% in this study, it is expected that the detection rate will rise to 81.6 % and 96.2 % for the sample size 3,000 and 8,000, respectively (Fig. 1d). Since most of very-rare variants likely arose within 5,000–10,000 years[10], they are essentially population-specific. Hence, most of them would be difficult to be imputed from a reference panel that was constructed from diverse genetic backgrounds. Indeed, we showed here significantly improved imputation accuracy from 1KJPN data when compared with 1KGP (Fig. 4 and Supplementary Fig. 7).

In 1KJPN, we found the signature of purifying selection on SNVs in regulatory elements located in both coding and noncoding sequences. This observation illustrates that this SNV set is expected to contain many very-rare variants that can be associated with diseases, and thus should be useful in future GWASs to fully capture causal variants.

This high-coverage strategy has also defined fine characteristics of SVs, CNVs and HLA types in a population scale. Although a substantial number of novel SVs was discovered by the high-coverage sequencing, we still had difficulty in discovering insertions longer than 100 bp because of the limitation of sequence read lengths (162 bp paired-end). In addition, simple repeat regions and structurally complex regions (for example, centromeric and telomeric sequences) were not covered in our SV discovery. To build a more comprehensive reference panel with SVs, alternative approaches, such as the use of longer sequence reads, would be necessary.

Since the cost of WGS is gradually decreasing, population-wide sequencing, as in the present study, will become a reasonable

approach to discover the full spectrum of SNVs, SVs and CNVs in a reliable manner. However, it remains challenging to sequence more than tens of thousands of samples in cohort projects even with the sophistication of current sequencing technologies. Thus, a hybrid strategy that combines both sequencing and genotyping with a customized SNP array optimized to a specific population background, for example, Japonica Array[58], becomes highly desirable. Our hope is that 1KJPN will foster basic research by amalgamating more accurate genotype imputation in cohort studies with medical information, and thereby aid in constructing an advanced medical system to improve the quality of health-care services.

Methods

The Supplementary Note in this manuscript lists details of samples, data generation, selection protocols with the SNP array and also informs the bioinformatics not covered in this section, for example, the estimation of HLA types, imputation and simulation of population genetics.

Sample information. This project was performed as part of the prospective cohort study at the ToMMo, with the approval of the ethical committee of the Tohoku University School of Medicine. The samples used here were from the cohort participants, all of whom gave their written consent. In our informed consent to the participants to our cohort project, whole-genome data including 'sequenced data, variant calls and inferred genotypes' are securely controlled under the Materials and Information Distribution Review Committee of Tohoku Medical Megabank Project and the data sharing to researchers are discussed in each research proposal by the review committee.

The current status of our committee is to put the part of the SNP frequency data as open data in the National Bioscience Database Center website (http://humandbs.biosciencedbc.jp/en/). Every year, the data availability policy will be updated with the review committee.

Whole-genome sequencing. To prevent sample mix-up, samples were handled in 96-well plates during the course of the library construction. The genomic DNA samples were diluted to a concentration of $10\,ng\,\mu l^{-1}$ using a laboratory automation system (Biomek NXP, Beckman), and fragmented using a 96-well plate using the DNA sonication system (Covaris LE220, Covaris) to a target size of 550 bp, on average. The sheared DNA was subjected to library construction with the TruSeq DNA PCR-Free HT sample prep kit (Illumina) using a Bravo liquid-handling instrument (Agilent Technologies). Finally, the completed libraries were transferred into 1.5 ml tubes that were labelled with a barcode, and were then denatured and neutralized, and processed for library QC.

The library quantitation and QC were performed with quantitative MiSeq (qMiSeq), a newly developed method for library quantitation. In this protocol, 8 or 10 µl of prepared libraries was denatured with an equal volume of 0.1 N NaOH for 5 min at room temperature, and diluted with a 49-fold volume of ice-cold Illumina HT1 buffer. Overall, 50 µl of 96 denatured libraries including three control samples (those examined earlier on HiSeq) were pooled. Then, 60 µl of the pooled library was diluted with 540 µl of ice-cold Illumina HT1 buffer and analysed with MiSeq using a 25-bp, paired-end protocol. We utilized the index ratio determined by the MiSeq sequencer, as a relative concentration to determine the run condition for the HiSeq run. To share the details of qMiSeq, we have published a methodology paper[27]. In addition to qMiSeq, an electrophoretic analysis using the Fragment Analyzer (Advanced Analytical) software was performed as a part of the library QC.

DNA libraries were analysed using the HiSeq 2500-sequencing system, according to the manufacturer's protocol. A TruSeq Rapid PE Cluster Kit (Illumina), and one-and-a-half TruSeq Rapid SBS Kit (200 cycles, Illumina), were used to perform a 162-bp, paired-end read in Rapid-Run Mode. On the basis of the qMiSeq results, the libraries had been diluted to the appropriate concentrations and were used for on-board cluster generation (Illumina). We routinely checked the cluster density at the first base report and decided whether to continue the analysis depending on the density ($\sim$550–650 K mm^{-2}).

SNV step 1 alignment and variant call for each sample. The sequence reads from each sample were aligned to the reference human genome (GRCh37/hg19) with the decoy sequence (hs37d5). Two aligners, Bowtie2 (version 2.1.0) with the '-X 2000' option, and BWA-MEM (ver. 0.7.5a-r405) with the default option, were used. To call the SNVs, the Bcftools software (ver. 0.1.17-dev) and the Genome Analysis Toolkit (GATK version 2.5-2) were, respectively, applied to each aligned result (Step 1 in Supplementary Fig. 2). For each sample, the read depth of each SNV position was calculated for the downstream filtering steps (Supplementary Fig. 2; middle-bottom in Step 1). Here the read depth represents the total number of sequence reads aligned to the SNV position, with the mapping quality being more than or equal to 5.

We compared the SNV calls and the genotyping results from the HumanOmni2.5-8 BeadChip for the same samples to evaluate the precision and the power (= recall) at designed positions in the SNP array. The precision, as calculated using Bowtie2 with the Bcftools variant caller (Bowtie2 + Bcftools), was always better than that obtained using BWA-MEM with GATK. On the contrary, the power obtained using BWA-MEM + GATK was always better than that obtained with Bowtie2 + Bcftools (Supplementary Data 2). These results are consistent with a previously conducted analysis[59]. Thus, the SNVs obtained using BWA – MEM + GATK were considered the result of highly sensitive SNV detection (trying to find as many SNVs as possible), whereas the SNVs that were obtained using bowtie2 + Bcftools were used as the candidate of the high-confidence SNVs to apply to the downstream filtering steps. Hereafter, the term 'SNVs' will refer to variants called using Bowtie2 + bcftools.

SNV step 2 genotype depth filter for each individual. We filtered out the genotype calls that had an extraordinarily low or high read depth. For each sample, we determined the range of the read depth on the SNVs to be retained, according to the precision and the recall from the comparison with the genotyping results of the HumanOmni2.5-8 BeadChip. The SNVs were grouped by read depth of next generation sequencing (NGS) (we only used the sequenced reads with a mapping quality more than or equal to 5, to avoid the reads being aligned to multiple chromosomal locations), and then precision and recall were calculated for each group (Fig. 1a and Supplementary Fig. 2; the middle plot in Step 2). The SNVs in the groups with a precision value > 0.998 were extracted as reliable SNVs (Supplementary Fig. 2 SNV Filter; the left plot and figure in Step 2). In this step, 2.0% of the segregating sites were filtered out (Step 2 in Supplementary Table 1).

SNV step 3 SNV depth filter for each locus. We then filtered out the SNV loci in which more than 10 % of the samples were not genotyped because of the depth filter applied in the previous step. For example, the SNV calls in such loci might suffer from an extraordinarily high or low depth, because of the read alignment to repetitive sequences. The pass ratio of the SNV sites was calculated (Supplementary Fig. 2; the middle figure in Step 3). In this step, 4.99% of the SNV sites were filtered out (Step 3 in Supplementary Table 1).

SNV step 4 genome complexity filter. It is difficult to call SNVs in low-complexity genomic regions, such as short tandem repeats. The RepeatMasker programme annotated 56% of genomic regions as low-complexity regions. The precision of each annotated group was calculated and compared with the genotyped SNP array of the same sample (Supplementary Fig. 2; the left table in Step 4). The precision of the SNV calls was calculated for each genomic region by comparing them with the genotype calls obtained with the HumanOmni2.5-8 Bead-Chip. Then, the SNVs in the genomic regions whose precision was $\leq$0.997 were removed (Supplementary Fig. 2; the right figure in Step 4), that is, Alu, ERVK, Low_complexity, Satellite, Simple_repeat, TcMar-Mariner and others (more detail in Supplementary Data 3). In this step, 14.22% of SNVs were filtered out (Step 4 in Supplementary Table 1).

SNV step 5 tool bias filter. To control the bias of the tool used in Step 1, we only retained the SNV sites that were also discovered using the other method. Thus, we removed the SNVs that were not detected by BWA-MEM + GATK, that is, the highly sensitive SNVs. At the top of the figure in Step 5 (Supplementary Fig. 2), the leftmost variant was not detected using the other tool, and thus those SNVs were not used in the downstream analysis. In this step, 0.57% of the SNVs were filtered out (Step 5 in Supplementary Table 1).

SNV step 6 population genetics filter. Finally, we filtered out the SNVs with a HWE test P value$< 10^{-5}$, using the genotype frequencies of the SNVs determined in Step 5. This filter removed the SNV-genotype frequencies that are not consistent with the HWE. Most of the SNVs that were filtered out might be artefacts from the incompleteness of the reference genome, or from systematic alignment errors. In this step, 1.03% of the SNVs were filtered out, and in the end, 77.20% of the raw SNVs were mined through these filtering steps (Step 6 in Supplementary Table 1).

SNV validation. The autosomal SNVs with a MAF$\leq$0.5%, a MAF of 0.5–1%, a MAF of 1–5% and a MAF > 5% were subjected to validation. For each frequency, the MassARRAY design (Sequenom) was applied to 1,000 potential variants called in at least one of the 12 representative samples, which were randomly selected from 1,070 samples (in the MassArray experiments, as maximum 36 SNVs can be evaluated at once. There are some limitations for the combination of SNVs. Thus, we have started from enough 1,000 SNV candidates in the SNV validation). Among the SNPs that passed an assay design step (809, 813, 832 and 819 sites for each frequency, respectively), 322 sites (108, 70, 72 and 72 sites for each frequency, respectively) were selected and analysed using the Sequenom MassARRAY (Sequenom). The same sites were also analysed using amplicon-sequencing.

The MassARRAY analyses were performed according to the manufacturer's protocol. The design of the PCR and the extension primers was performed using the Assay Design suite (Sequenom). The genomic regions containing potential variants were amplified by multiplex PCR (35 or 36 sites for each region). The single-base extension reaction was followed by alkaline phosphatase treatment, and reaction mixtures were purified with resin and spotted on 96-well SpectroCHIPs using a Nanodispenser. SpectroCHIPs were analysed using the MassARRAY system. Variant calls were performed using the SpectroTyper software (ver. 4.0). The Sequenom validation was applied to 94 samples (including the 12 validation samples). If a variant was confirmed in at least one of the twelve validation samples, the variant was marked as being true. In the event that the call rate was less than 80%, or that a nonspecific, single-base extension was observed in the no-template control, the assay was considered unreliable and was marked as a no-call.

The genomic regions examined using the MassARRAY were amplified using a two-step PCR to add Illumina sequencing adaptors. In total, 322 regions of the 12 samples were sequenced with 76 bp, paired-end reads using a MiSeq sequencer with the MiSeq Reagent Kit v3. Index sequences (Illumina D701-D712) were used to distinguish the 12 validation samples. To validate the SNVs, we performed the following data analysis. (1) For each SNV site, the nucleotide sequences comprising both the 300 bp upstream and downstream of the reference allele, and of the alternative allele, were prepared, (2) the sequence reads were aligned to these reference sequences using BWA-MEM[60] and (3) based on the depth of coverage at each site, the SNV call was validated. The regions of low coverage (less than 30 ×) were marked as a no-call. If a variant was confirmed in at least one of the 12 validation samples, it was marked as being true.

Sanger sequencing was performed using the BigDye terminator cycle sequencing kit, v.3.1 (Applied Biosystems), in accordance with the manufacturer's instructions. The SNVs that had discordant genotype calls between the HiSeq and the MassARRAY analyses, or between the HiSeq and the amplicon-sequencing analyses, were further validated using Sanger sequencing. The genomic DNA fragments containing one of the SNVs requiring validation were amplified by PCR, and these fragments were then sequenced in both the 5' and the 3' directions. The oligonucleotide sequences used for PCR and for sequencing are available on request.

From the validated 322 sites, 282 sites were remained in the high-confidence SNVs set. Finally, 62, 124 and 96 SNVs were the validation targets in rare and very-rare, low and common MAF class, respectively. If either the MassARRAY or the amplicon-sequencing analysis turned out to yield a no-call, the SNV site was also treated as a no-call. The SNV sites that were validated by both methods were treated as true positives. If the two methods produced inconsistent genotypes, we assumed that the genotypes determined by Sanger sequencing are true. The FDR was calculated by dividing the number of false variants by the sum of the true and false variants (Supplementary Table 2).

To validate the newly discovered SNVs in this study, we also conducted the validation experiment using a custom Illumina BeadChip array. The custom array contained probes of 25,317 novel SNVs that were arbitrarily picked from the novel variants in the high-confidence SNV set. These SNVs include SNVs in intergenic regions as well as in coding regions, some of which were of biological interest (for example, nonsynonymous or stop-gained variants).

The genotyping of 854 samples, which are part of the 1KJPN, was conducted using this custom array with the same protocol used for the HumanOmni2.5-8 BeadChip genotyping described above. We validated the SNV discovery at the designed sites, and calculated per site FDR of novel SNVs. If a SNV is discovered in several samples and more than one samples could detect the same SNV among these samples by using the SNP array, then the SNV is considered to be correctly discovered and the SNV is counted as true. Otherwise, the SNV is counted as false. Finally, the total FDR for each category (common, low, rare and very-rare) in Supplementary Table 3 was calculated as follows: overall per site FDR = (FDR of known SNV) × f_{known} + (FDR of novel SNV) × (1 − f_{known}), where f_{known} is the fraction of SNVs listed in the single nucleotide polymorphism database (dbSNP) and FDR of known SNV as 0.

Indel validation. Autosomal indels ranging in size from 1 to 30 bp were subjected to validation. The 150 potential indels called in at least one of the twelve representative samples, which were randomly selected from the total 1KJPN samples, were applied to the MassARRAY design (Sequenom). Among the 98 indels that passed the assay design step, 95 sites were randomly selected and analysed using amplicon sequencing.

The 95 regions of the 12 samples were sequenced with 76-bp, paired-end reads using a MiSeq sequencer with the MiSeq Reagent Kit, v3. Index sequences (Illumina D701-D712) were used to distinguish between the 12 validation samples. To validate the indel calls, we performed the following data analysis: (i) for each indel site, the nucleotide sequences comprising both the 300 bp upstream and downstream of the reference allele, and of the alternative allele (either insertions or deletions), were prepared, (ii) the sequence reads were aligned to these reference sequences using BWA-MEM and (iii) based on the depth of coverage at each site, the indel call was validated. The regions of low coverage (less than 100 ×) were marked as no-calls. If an indel was confirmed in at least one of the twelve validation samples, it was marked as being true.

If the amplicon-sequencing analysis turned out to yield a no-call, we tried to determine the genotypes using Sanger sequencing. If Sanger sequencing could not determine the genotypes, the variants were treated as a no-calls. The FDR was calculated by dividing the number of false variants by the sum of the true and false variants (Supplementary Table 2).

Variant annotation. The effects of the coding and the intergenic variants were predicted by the SnpEff software (ver. 3.3c), based on the gene annotation model of GENCODE version 17. The functional categories of synonymous, nonsynonymous, intron, 5'-UTR and 3'-UTR were applied to the SNVs whose transcript label was 'protein-coding'. For the insertion and the deletion variants, the variants on the protein-coding region were labelled according to the impact record of the SnpEff annotation. The category of 'intergenic' was applied for intergenic regions when no annotation was assigned to the transcript label. The functional consequences of nonsynonymous SNVs were retrieved from the PolyPhen-2 records of the dbNSFP 2.0 database[61]. The scaled Combined Annotation Dependent Depletion scores (C scores)[34] were added to each SNV by intersecting the list of the precomputed C scores with all possible SNVs. The predictions of the chromatin state were obtained from the ENCODE Project. The predictions were categorized into seven chromatin states—TSS: predicted promoter region including TSS; PF: promoter-flanking region; E: enhancer; WE: weak enhancer or open chromatin *cis*-regulatory element; CTCF: CTCF-enriched element; T: transcribed region; and R: repressed or low-activity region. The combined predictions of ChromHMM[33] and Segway segmentation[62] in the GM12878 cell line were assigned to the respective SNVs. The RNA regions (miRNA, lincRNA (long intergenic non-coding RNA), small nuclear RNA (snRNA) and snoRNA (small nucleolar RNA)) of the SNVs were annotated from several sources. The annotations of miRNA and lincRNA were performed according to the miRBase[63] and the Human lincRNA Catalog[64], respectively. Only the mature parts of the miRNAs were used for annotation (Supplementary Fig. 4). The annotation model of GENCODE17 was used for annotation of the snRNAs and of the snoRNAs.

Site frequency spectrum. The SFSs were analysed to assess the impact of natural selection on each category of SNVs. The number of genotyped individuals to obtain the high-confidence SNVs varies from one SNV to another because of the individual depth filter (Step 2 in Supplementary Fig. 2). Thus, we applied the hyper-geometric projection[65], in which each variant was downsampled to a uniform sample size across all SNV sites. The downsampled size was 963, which was the minimum number of genotyped individuals to obtain the high-confidence SNVs. To directly compare the SFSs in the 1KGP with those in the 1KJPN, we also downsampled the SNVs of the 1KGP to the uniform sample size of 963. The FVRV was calculated by dividing the number of very-rare SNVs by the total number of SNVs within the resampled SFS.

Structural variants. SVs were called using the BreakDancer[66] software (ver. 1.1), the Pindel[67] software (ver. 0.2.5a3) and the Haplotype Caller software implemented in the Genome Analysis Toolkit[68] (version 2.5–2), from the binary of sequence alignment/map (BAM) files constructed with BWA-MEM (version 0.7.5a-r405). For the insertions and deletions of a length < 100 bp, calls from Haplotype Caller were used. For the deletions of a length between 100 bp and 1 M bp, calls from Pindel (≥ 100 bp) and BreakDancer (≥ 1 kbp) were used, and merged into one unified call set, as described in the integrated structural variation pipeline[35]. SV calls that were overlapping by ≥ 80% were merged into one unified call. The start and end positions of the merged call were defined as the mean of the start and the end positions of the original calls, respectively. The novelty rates of SV calls were calculated based on the dbSNP 138 database and the Database of Genomic Variants, as of 23 July 2013. The insertion calls whose nearest known insertion was within 10 bp, and the deletion calls that overlapped by ≥ 50% with a known deletion, were regarded as known calls.

Whole-genome identification of genic CNVs. CNVs were called with the CNVnator[69] software (ver. 0.3), using default options and a 100-bp bin size. We filtered out calls when 90% of the bases of the called region were 'N' in the hg19 reference genome, or if 'cnvnatorP1' or 'cnvnatorP2' was ≥ 1. The CNV calls that overlapped by ≥ 80% were merged into one unified call. The start and the end positions of the merged calls were defined as the mean of the start and the end positions of the original calls, respectively. The autosomal CNVs whose biallelic MAF was ≥ 1%, length was ≥ 1,000 bp and positions were overlapping with a human gene annotation in Ensembl (release-75) were analysed. The copy numbers of each CNV call in the 1KJPN that overlapped with human gene loci are summarized in Supplementary Data 3.

Diploid copy-number estimation and validation of the AMY1 genes. First, we counted the number of reads that aligned to the following regions: AMY2B (chr1:104096436–104122156), AMY1A (chr1:104,096,436–104,207,173), AMY1B (chr1:104,230,036–104,239,302), AMY1C (chr1:104,293,027–104,301,312), Region X (chr1:104,210,000–104,211,500), Region Y (chr1:104,350,000–104,400,000) and Region Z (chr1:104,400,000–104,450,000), where the genomic coordinates (hg19

reference genome) are illustrated (see Fig. 3c and Supplementary Fig. 6a,b). Region Z was chosen on the basis of the mappability of the reference genome so that the depth of coverage in each region across the whole genome in each sample can be determined on the basis of the average depth of this region. We then normalized read counts in these regions, by dividing them by half of the counts in Region Z, so that the normalized counts became two in diploid regions. The sum of the normalized counts in *AMY1A*, *AMY1B* and *AMY1C*, and the normalized counts in Region X, were calculated for each sample (Fig. 3c). Region X is located between the *AMY1A* and *AMY1B* regions, and most of the reads were uniquely aligned to this region. Three *AMY1* gene loci (*AMY1A*, *AMY1B* and *AMY1C*) in the reference genome are very similar to each other (having more than 99.9% sequence identity), and hence it is difficult to distinguish one locus from another. The normalized coverages in *AMY1* and in Region X were highly correlated (Fig. 3c), which suggests that the locus that contains *AMY1A* and *AMY1B* is a copy unit of the salivary amylase gene. While all samples had two diploid copies of Region Y, 17 samples had more than two diploid copies of *AMY2B*.

Digital PCR was performed with the QuantStudio 3D Digital PCR System (Life Technologies) according to the manufacturer's protocol. Diploid copy numbers of the *AMY1* gene in 10 samples (whose diploid copy numbers were predicted to be six, seven or eight) were determined using predesigned TaqMan Copy Number Assay (Hs07226362_cn for *AMY1A*, *AMY1B* and *AMY1C*, Life Technologies) and normalizing with TaqMan Copy Number Reference Assay RNase P (Life Technologies). The PCR reaction mixture contained 12 ng of MspI-digested genomic DNA, 7.5 µl of QS3D digital PCR Master Mix (Life Technologies), 0.75 µl of TaqMan Copy Number Assay and 0.75 µl of TaqMan Copy Number Reference Assay in a final volume of 15 µl. The data were analysed with QuantStudio 3D AnalysisSuite (Life Technologies), and the quantification of the target genomic region was presented as the number of copies per microlitre of PCR mixture. The estimated diploid copy numbers of *AMY1* with WGS were consistent with the result obtained with the digital PCR analysis (Supplementary Fig. 5).

References

1. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
3. Ozaki, K. *et al.* Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat. Genet.* **32**, 650–654 (2002).
4. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and. *Nature* **447**, 661–678 (2007).
5. Reich, D. E. & Lander, E. S. On the allelic spectrum of human disease. *Trends Genet.* **17**, 502–510 (2001).
6. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
7. Almgren, P. *et al.* Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. *Diabetologia* **54**, 2811–2819 (2011).
8. Jiao, S. *et al.* Estimating the heritability of colorectal cancer. *Hum. Mol. Genet.* **23**, 3898–3905 (2014).
9. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
10. Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216–220 (2013).
11. Nelson, M. R. *et al.* An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* **337**, 100–104 (2012).
12. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
13. Altshuler, D. M. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
14. Scherer, S. W. *et al.* Challenges and standards in integrating surveys of structural variation. *Nat. Genet.* **39**, S7–S15 (2007).
15. Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368–372 (2010).
16. Stefansson, H. *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232–236 (2008).
17. McCarroll, S. A. *et al.* Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nat. Genet.* **40**, 1107–1112 (2008).
18. Perry, G. H. *et al.* Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* **39**, 1256–1260 (2007).
19. Valsesia, A., Macé, A., Jacquemont, S., Beckmann, J. S. & Kutalik, Z. The growing importance of CNVs: new insights for detection and clinical interpretation. *Front. Genet.* **4**, 92–92 (2013).
20. The Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat. Genet.* **46**, 818–825 (2014).
21. Carmi, S. *et al.* Sequencing an Ashkenazi reference panel supports population-targeted personal genomics and illuminates Jewish and European origins. *Nat. Commun.* **5**, 4835 (2014).
22. Wong, L. P. *et al.* Deep whole-genome sequencing of 100 southeast Asian Malays. *Am. J. Hum. Genet.* **92**, 52–66 (2013).
23. Gudbjartsson, D. F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47**, 435–444 (2015).
24. Sims, D., Sudbery, I., Ilott, N. E., Heger, A. & Ponting, C. P. Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* **15**, 121–132 (2014).
25. Aird, D. *et al.* Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* **12**, R18 (2011).
26. Liang, W. S. *et al.* Long insert whole genome sequencing for copy number variant and translocation detection. *Nucleic Acids Res.* **42**, e8 (2014).
27. Katsuoka, F. *et al.* An efficient quantitation method of next-generation sequencing libraries by using MiSeq sequencer. *Anal. Biochem.* **466**, 27–29 (2014).
28. Sato, Y. *et al.* SUGAR: graphical user interface-based data refiner for high-throughput DNA sequencing. *BMC Genomics* **15**, 664 (2014).
29. Williamson, S. H. *et al.* Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proc. Natl Acad. Sci. USA* **102**, 7882–7887 (2005).
30. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695 (2009).
31. Cooper, D. N., Ball, E. V. & Krawczak, M. The human gene mutation database. *Nucleic Acids Res.* **26**, 285–287 (1998).
32. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
33. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
34. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
35. Mimori, T. *et al.* iSVP: an integrated structural variant calling pipeline from high-throughput sequencing data. *BMC Syst. Biol.* **7 Suppl 6**, S8 (2013).
36. Groot, P. C. *et al.* The human alpha-amylase multigene family consists of haplotypes with variable numbers of genes. *Genomics* **5**, 29–42 (1989).
37. Gilissen, C. *et al.* Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511**, 344–347 (2014).
38. Stefansson, H. *et al.* CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature* **505**, 361–366 (2014).
39. de Bakker, P. I. W. *et al.* A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat. Genet.* **38**, 1166–1172 (2006).
40. Nariai, N. *et al.* HLA-VBSeq: accurate HLA typing at full resolution from whole-genome sequencing data. *BMC Genomics* **16 Suppl 2**, S7 (2015).
41. Robinson, J. *et al.* The IMGT/HLA database. *Nucleic Acids Res.* **41**, D1222–D1227 (2013).
42. Itoh, Y. *et al.* High-throughput DNA typing of HLA-A, -B, -C, and -DRB1 loci by a PCR-SSOP-Luminex method in the Japanese population. *Immunogenetics* **57**, 717–729 (2005).
43. Marks, C. Immunobiological determinants in organ transplantation. *Ann. R Coll. Surg. Engl.* **65**, 139–144 (1983).
44. Shiina, T., Hosomichi, K., Inoko, H. & Kulski, J. K. The HLA genomic loci map: expression, interaction, diversity and disease. *J. Hum. Genet.* **54**, 15–39 (2009).
45. Noguchi, E. *et al.* Genome-wide association study identifies HLA-DP as a susceptibility gene for pediatric asthma in Asian populations. *PLoS Genet.* **7**, e1002170 (2011).
46. Dahi, P. B. *et al.* Donor-recipient allele-level HLA matching of unrelated cord blood units reveals high degrees of mismatch and alters graft selection. *Bone Marrow Transplant.* **49**, 1184–1186 (2014).
47. Kojima, K. *et al.* HapMonster: a statistically unified approach for variant calling and haplotyping based on phase-informative reads. *Lect. Notes Comput. Sci.* **8542**, 107–118 (2014).
48. Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
49. Kamada, F. *et al.* A genome-wide association study identifies RNF213 as the first Moyamoya disease gene. *J. Hum. Genet.* **56**, 34–40 (2011).
50. Liu, W. *et al.* Identification of RNF213 as a susceptibility gene for moyamoya disease and its possible role in vascular development. *PLoS ONE* **6**, e22542 (2011).
51. Lim, E. T. *et al.* Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet.* **10**, e1004494 (2014).
52. Feuchtbaum, L., Carter, J., Dowray, S., Currier, R. J. & Lorey, F. Birth prevalence of disorders detectable through newborn screening by race/ethnicity. *Genet. Med.* **14**, 937–945 (2012).
53. Yamaguchi, S. Newborn screening in Japan: restructuring for the new era. *Ann. Acad. Med. Singapore* **37**, 13–15 (2008).

54. Mohr, J. Estimation of linkage between the Lutheran and the Lewis blood groups. *Acta Pathol. Microbiol. Scand.* **29,** 339–344 (1951).

55. Kudo, T. *et al.* Molecular genetic analysis of the human Lewis histo-blood group system. II. Secretor gene inactivation by a novel single missense mutation A385T in Japanese nonsecretor individuals. *J. Biol. Chem.* **271,** 9830–9837 (1996).

56. He, M. *et al.* A genome wide association study of genetic loci that influence tumour biomarkers cancer antigen 19-9, carcinoembryonic antigen and alpha fetoprotein and their associations with cancer risk. *Gut* **63,** 143–151 (2014).

57. Lin, X. *et al.* Genome-wide association study identifies novel loci associated with serum level of vitamin B12 in Chinese men. *Hum. Mol. Genet.* **21,** 2610–2617 (2012).

58. Kawai, Y. *et al.* Japonica array: improved genotype imputation by designing a population-specific SNP array with 1070 Japanese individuals. *J. Hum. Genet.* (doi: 10.1038/jhg.2015.68) (2015).

59. Li, H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30,** 2843–2851 (2014).

60. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).

61. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum. Mutat.* **34,** E2393–E2402 (2013).

62. Hoffman, M. M. *et al.* Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat. Methods* **9,** 473–476 (2012).

63. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42,** D68–D73 (2014).

64. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25,** 1915–1927 (2011).

65. Nielsen, R., Hubisz, M. J. & Clark, A. G. Reconstituting the frequency spectrum of ascertained single-nucleotide polymorphism data. *Genetics* **168,** 2373–2382 (2004).

66. Chen, K. *et al.* BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat. Methods* **6,** 677–681 (2009).

67. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25,** 2865–2871 (2009).

68. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43,** 491–498 (2011).

69. Abyzov, A., Urban, A. E., Snyder, M. & Gerstein, M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* **21,** 974–984 (2011).

Author contributions

S.K., A.H. and N.Y. organized the cohort. H.K. and N.Y. organized the medical support team in the disaster area. Ni.M. organized the biobank. J.Y. organized the group of sequence analysis. NGS data were mainly generated and qualified by K.T., X.P, S.I., J.Yo. and F.K. SNP array and custom array experiments were generated by S.N., R.S. and I.D. Validations of SNVs, indels and CNVs were conducted by J.Y., S.S., S.I., K.T., Y.K. and F.K. M.N. organized the group of *in silico* analysis. K.Ki. and M.N. designed and implemented the super computer. Y.S., M.N., N.N. and K.K. checked the quality of the generated sequenced data. M.N., N.N. and K.K. mainly developed the NGS analysis pipeline on the super computer. Y.Y.-K., I.D., S.N., R.S. and M.N. chose the reference panel candidates by removing closed individuals and outliers. M.N., K.K., N.N., T.M. and Y.K. performed SNV analysis. Y.K. performed demography and very-rare variant functional analysis. N.N. and T.M. performed SV, CNV and HLA analyses. K.K. and M.N. performed imputation and GWAS analysis. Y.Y.-K. performed the annotation of pathogenic variants and the calculation of individual variant load. S.K., N.F., O.T. and J.D.E. provided the critical feedback on the manuscript. M.N., J.Y., F.K., N.N., K.K., Y.K., Y.Y.-K. and M.Y. wrote the manuscript. All authors including ToMMo Japanese Reference Panel Project have seen and approved the final manuscript.

ToMMo Japanese Reference Panel Project

Akito Tsuboi[1,7], Fuji Nagami[1,2], Hiroshi Kawame[1,2], Hiroaki Tomita[1,2,4], Ichiro Tsuji[1,2], Jun Nakaya[1,2], Junichi Sugawara[1,2], Kichiya Suzuki[1,2], Masahiro Kikuya[1,2], Michiaki Abe[1,2], Naoki Nakaya[1,2], Noriko Osumi[1,2], Riu Yamashita[1,2,3], Soichi Ogishima[1,2], Takako Takai[1,2], Teiji Tominaga[1,2], Yasuyuki Taki[1,†], Yoichi Suzuki.[1,2]

[7]Graduate School of Dentistry, Tohoku University, 4-1 Seiryo-machi, Aoba-ku, Sendai 980-8575, Japan.

症例報告

先天性門脈体循環短絡症に対して
コイル塞栓術を施行し得た3歳女児例

佐藤信一[1)], 木村正人[1)], 川野研悟[1)], 和田 基[2)], 高瀬 圭[3)], 坂本 修[1)], 呉 繁夫[1)]

東北大学医学部 小児科[1)], 同 小児外科[2)], 同 放射線診断科[3)]

Successful transcatheter closure of a congenital portsystemic shunt with interlocking detachable coils in a 3-year-old girl

Shinichi Sato[1)], Masato Kimura[1)], Kengo Kawano[1)], Motoshi Wada[2)]
Kei Takase[3)], Osamu Sakamoto[1)], Shigeo Kure[1)]

Department of Pediatrics[1)] and Pediatric Surgery[2)] and Diagnostic Radiology[3)], Tohoku University School of Medicine

Abstract Congenital portsystemic shunts (CPSS) are rare vascular anomalies in which the portal vein drains into a systemic vein, bypassing the liver through a partial or complete shunt. The clinical expression of CPSS ranges from asymptomatic to the development of many complications such as hepatic encephalopathy, mental retardation, hepatopulmonary syndrome, pulmonary hypertension, and hepatic tumor. For symptomatic CPSS, surgical ligation or transcatheter embolization should be undertaken with caution for portal hypertension.

We report a 3-year-old girl who presented hypergalactosemia in mass newborn screening. Contrast enhanced abdominal CT scan showed an extrahepatic portsystemic shunt, which connected to the portal vein via the inferior vena cava. Elevated plasma concentration of total bile acids (TBA) continued for three years and she developed dysarthria. She was successfully treated by transcatheter embolization with interlocking detachable coils. After embolization, plasma TBA fell to normal and the high signal on T1-weighted image in globus pallidus on brain MRI disappeared. Transcatheter embolization is effective for the treatment of extrahepatic portsystemic shunt. Early detection and appropriate management are important to achieve a good outcome.

Keywords Congenital portsystemic shunts, Transcatheter embolization, Hypergalactosemia, Mass newborn screening

症 例

3歳9か月女児
主訴：高ガラクトース血症

出生歴：在胎32週5日，切迫早産のため2絨毛膜2羊膜双胎I児として1,450gで帝王切開にて出生した．前医NICUにて出生後，呼吸障害や黄疸のためそれぞれ酸素投与や光線療法で加療され，日
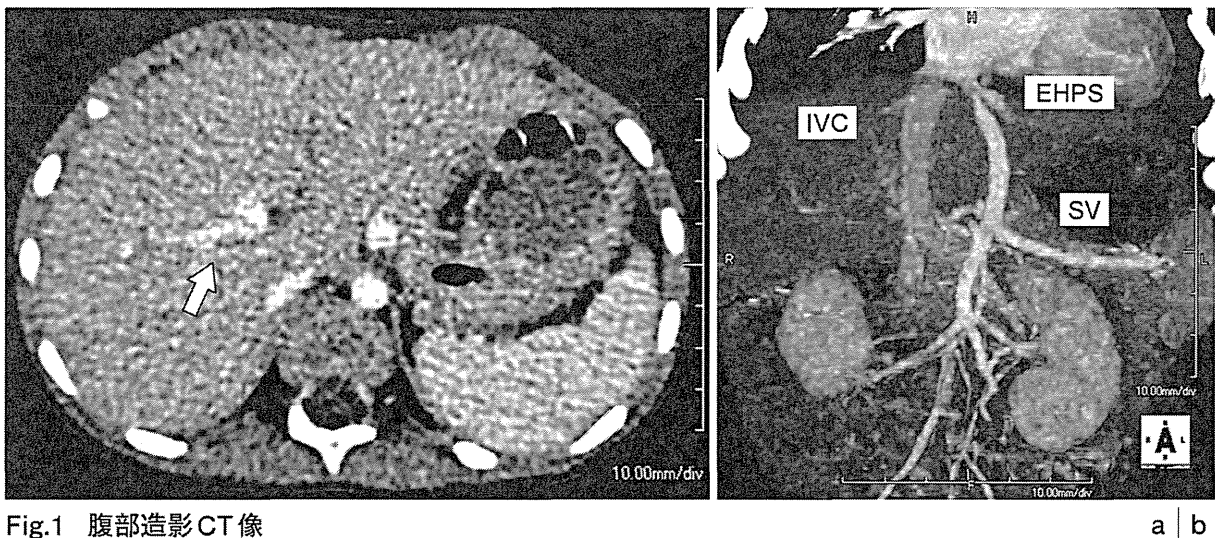
54

Fig.1　腹部造影CT像　　　　　　　　　　　　　　　　　　　　　　　　　　a｜b
　　　a：腹部造影CT横断像にて肝内門脈の描出を認める.
　　　b：造影CT冠状断MPR（multi-planar reformation）像において肝外門脈体循環短絡
　　　　（EHPS）は下大静脈ー右房接合部直近に合流しており合流部は狭窄している.
　　　　また，正常の門脈は描出されていないが，肝内に低形成な門脈様構造を認める.
　　　　IVC : inferior vena cava,　EHPS : extrahepatic portosystemic shunt
　　　　SV : splenic vein.

齢57に2,270gで退院した.

現病歴：日齢7に施行した初回の新生児マススク
リーニング検査において，ガラクトースは4mg/dℓ,
ボイトラー法で蛍光有りと正常であった. 体重
2,000g以上での再検（日齢48）でガラクトース
16mg/dℓと高値であり，さらに再々検査（日齢60,
67）においても16mg/dℓと高値が持続していた. 同
院で施行された腹部超音波検査では静脈管の開存
は指摘されなかった. 生後3か月（修正44週）で精
査目的に当科紹介となった. 当科初診時および再
診時（生後4か月）におけるガラクトースはそれぞ
れ2.3，3.6mg/dℓと基準値内であったが，生後5
か月時での血液検査で20.8mg/dℓと再び上昇を認
めた. 総胆汁酸は218.8μmol/ℓと高値であり，先
天性門脈体循環短絡症が疑われた. 腹部超音波検
査では静脈管開存症などの異常血管は検出されな
かったが，生後6か月に施行した腹部CT（Fig.1）
では門脈から分岐し下大静脈に流入する短絡血管
を認め先天性肝外門脈体循環短絡症と診断した.
ガラクトース制限食を導入し血中ガラクトースは
正常化したが，総胆汁酸は100〜150μmol/ℓと高
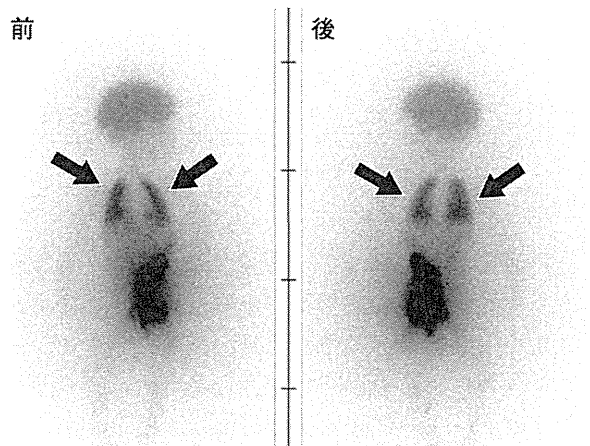値で持続した. 1歳7か月時に施行した[123]I-IMP
経直腸門脈血流シンチグラフィ（Fig.2）では門脈

Fig.2　[123]I-IMP経直腸門脈血流シンチグラフィー
　　　肝臓への集積は全体的に低く，両肺への集
　　　積が目立つ（矢印）. 投与65分後の像にお
　　　いて門脈体循環短絡率（肺の集積/肝臓の
　　　集積＋肺の集積）は76%と算出された.

から体循環へのシャント率は76％であり，頭部
MRI（Fig.3a）ではT1強調像で淡蒼球に高信号が
認められ，マンガンの沈着を疑った. 外科的処置
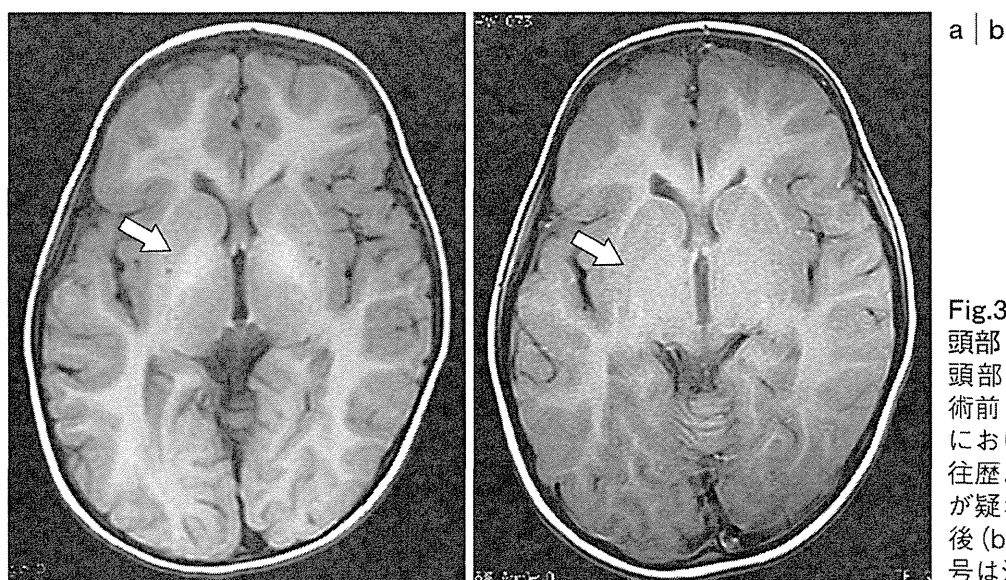につき当院小児外科にコンサルトしたところ，体
格が小さいこと（体重7kg），双胎同胞に比して明
らかな発達遅滞や神経症状が見られないことから

55

a | b

Fig.3
頭部 MR T1 強調像
頭部 MR T1 強調像にて
術前 (a) は淡蒼球 (矢印)
における高信号を認め既
往歴よりマンガンの沈着
が疑われたが, 閉鎖1年
後 (b) には同部位の高信
号は消失していた.

Table 1　入院時検査所見

WBC	9,300 /μℓ	TB	1.0 mg/dℓ	TP	6.2 g/dℓ
RBC	422×10⁴/mℓ	DB	0.1 mg/dℓ	Alb	3.7 g/dℓ
Hb	12.0 g/dℓ	AST	29 IU/ℓ	BUN	8.0 mg/dℓ
Plt	17.6×10⁴/mℓ	ALT	14 IU/ℓ	Cre	0.26 mg/dℓ
MCV	85.6 fl	ALP	849 IU/ℓ	NH3	72 μg/dℓ
MCH	28.5 pg	GTP	14 IU/ℓ	TBA	83.7 μmol/ℓ
MCHC	33.4%	ChE	288 IU/ℓ		

TBA：総胆汁酸

内科治療を継続することとなった. 3歳を過ぎた頃から双胎同胞に比べ明らかに発語の不明瞭さが目立つようになってきたため, 基底核へのマンガンの沈着による構音障害の進行を考慮し, 精査およびカテーテルによる閉鎖目的に入院となった.

入院時現症：身長 89 cm (-1.8 SD), 体重 11 kg (-1.5 SD). 発達異常はなかったが, 双子の妹と比較し構音障害を認めた. 呼吸音は清明, 胸骨左縁第二肋間に大動脈弁狭窄症に由来する Levine Ⅱ/Ⅵの収縮期雑音を聴取した. 腹部に異常は認めなかった.

入院時検査所見

血液検査 (Table 1) では, アンモニア 72 μg/dℓ, 総胆汁酸 83.7 μmol/ℓ と上昇を認めたが, その他検査値の異常は認めなかった.

腹部超音波検査では下大静脈と平行に走行し, その後下大静脈に合流する異常血管を認めた. 肝臓内では門脈主幹部の同定が困難であったが, 肝腫瘍や血管腫等の器質的異常は認められなかった.

入院後経過

右内頸静脈アプローチによりコブラ型 MOIYAN バルーンカテーテル (5 Fr, 80 cm, 最大バルーン径 11 mm, 東海メディカルプロダクツ, 名古屋) を用いて, 肝静脈圧及び肝静脈楔入圧, 次に下大静脈近位部から門脈への短絡血管へ挿入し, 上腸間膜静脈, 脾静脈, また短絡血管閉塞時の門脈圧を測定した. 閉塞試験では閉塞前の平均門脈圧は 17 mmHg であったが閉塞後7分間の観察で平均門脈圧は 20 mmHg までの上昇にとどまった. 腹部 CT 上も肝内門脈を認め, 血管撮影でも低形成ではあったが門脈構造を認めた. さらに, 短絡血管が下大静脈に合流する部分で狭窄しており, 脾静脈までの

56

距離があることからコイル塞栓の適応と判断した. 塞栓後のコイルが低形成の門脈起始部, 脾静脈にかからないようにIDC™ Interlocking Detachable Coil (ϕ14 mm, 20 cm, Boston Scientific, Natick, MA) をアンカーとして留置後, その内部にIDC™ Interlocking Detachable Coil (ϕ9 mm, 20 cm, Boston Scientific, Natick, MA) に加え Interlock™ Fibered IDC™ Occulusion System (ϕ6 mm, 10 cm, ϕ5 mm, 15 cm 各1本, Boston Scientific, Natick, MA) の計

4本のコイルで完全閉塞を得た. コイル塞栓後には上腸間膜動脈, 腹腔動脈, 脾動脈造影を行い, 短絡血管の血流が消失し上腸間膜静脈および脾静脈から門脈への血流は良好に保たれているのを確認した (Fig.4). コイル塞栓術の翌日に施行した腹部超音波検査では腹水などの門脈圧亢進症の症状は認めず, 短絡血管の血流の閉塞を確認した. 血液検査では, 総胆汁酸が術前 83.7 μmol/ℓ から術後 5.8 μmol/ℓ に低下しシャント血流が消失した

Fig.4 カテーテルによる短絡血管閉鎖術

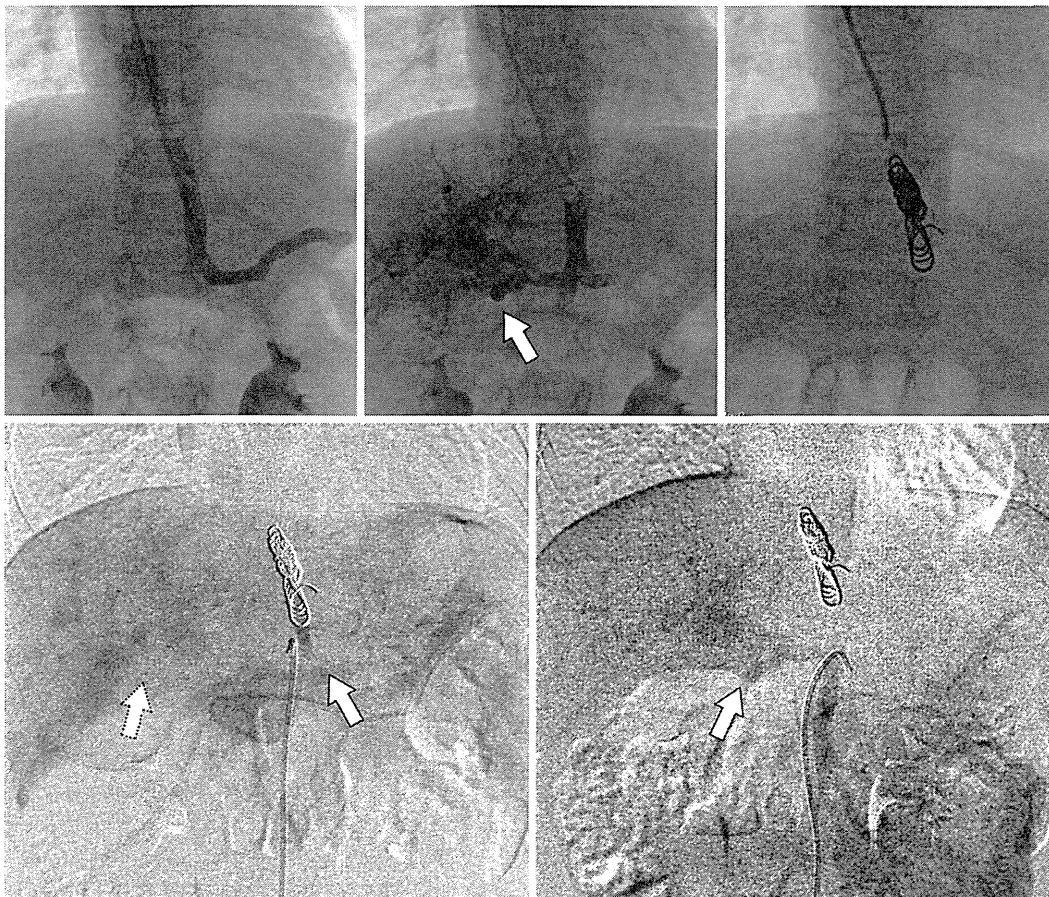| a	b	c
d		e

a : 門脈体循環短絡の造影により短絡血管は下大静脈の右房接合部付近に合流すること及び短絡血管が下大静脈合流部で狭窄していることが確認された.

b : バルーン閉塞下での門脈造影で門脈海綿状変形 (cavernous transformation) と低形成の門脈様構造を認めた.

c : IDC™ Interlocking Detachable Coil と Interlock™ Fibered IDC™ Occulusion System の計4本のコイルで完全閉塞を得た.

d : コイル塞栓後に施行した腹腔動脈造影の後期相で短絡血管の完全閉塞 (d, 実線矢印) と肝内門脈 (d, 点線矢印) の描出を確認した.

e : 上腸間膜動脈造影の後期相において門脈を確認した (d, e はデジタル・サブトラクション血管造影法による画像).

57

と考えられた．コイル塞栓術1年後においても構音障害に明らかな改善は認めなかったが，血中総胆汁酸値は正常化しており，頭部MRIではT1強調画像における淡蒼球の高信号はほぼ消失していた (Fig.3b)．

考　察

先天性門脈体循環短絡症は肝内門脈の有無により2つのタイプに分類され，肝内門脈を認めないTypeⅠ (先天性門脈欠損症) と肝内門脈を有するTypeⅡに分類される[1]．両者とも消化管由来の門脈血の一部または全てが肝臓を迂回して直接体循環に流入することが病態の主因となる．診断のきっかけは肝機能異常や肝腫瘤の精査によることもあるが，多くの場合には無症状であり，特に小児においては本症例のように新生児マススクリーニングで高ガラクトース血症を呈することが診断のきっかけになることが多い．先天性門脈体循環短絡症自体はまれな疾患であり，大阪市における10年間に施行された約24万件のスクリーニング検体のうち81件が高ガラクトース血症の要精査となり，そのうちの6件が門脈体循環短絡症として診断されている[2]．

先天性門脈体循環短絡症のうちTypeⅠの多くは肝移植の適応であるが，TypeⅡでは外科的またはカテーテルによる閉鎖術の適応となる．TypeⅡのうち静脈管は遅くとも生後3〜6か月までに自然閉鎖すると言われており，1〜2歳までに自然閉鎖が見られない場合には何らかの閉鎖術の適応を検討すべきであると考えられる[3]．また短絡率60％以上の症例や短絡の存在期間が長期になるほど肝性脳症の発症頻度が高くなる[4]ことから，年齢や短絡率を基に閉鎖術を行う時期を決定するのが望ましいと考えられる．本症例においては，頭部MRIにてマンガンの沈着を認めていること，不明瞭な発語が目立つことから短絡血管閉鎖術の方針となった．

閉鎖術の治療としては，軽症例に対する食餌療法などの保存的治療の他，根治的な短絡血管離断や結紮などの手術治療が行われてきたが，最近ではカテーテルによる血管内塞栓術による治療報告が増加してきている[5]．手術治療や血管内塞栓術のいずれの方法においても術前に肝内門脈の発達

評価を行い，門脈の閉塞試験を行うことで短絡血管閉塞後の合併症 (門脈圧亢進症) の予測を行う必要がある．この点に関して，Kamimatsuseらは閉塞試験において門脈圧が22mmHg以上に上昇すれば一期的閉鎖を断念すべき[6]との報告をしている．一方，Franchiらは32mmHg以下の場合には一期的閉鎖が可能であると述べている[7]．カテーテルによる血管内治療については手術よりも侵襲性が少なく，短絡血管を直接閉塞して門脈圧を測定しながら十分な門脈血流が確保できることを確認して治療を進めることができる利点がある．Shwartzら[8]が1999年に静脈管開存症に対して，Kimら[9]は2000年に先天性門脈体循環短絡症に対してコイルによる閉鎖術の有効性を報告している．さらに近年ではAmplatzer Vascular Plugにより有効な塞栓が得られたとする報告もなされている[10,11]．本症例においては閉塞試験において平均門脈圧は20mmHgの上昇にとどまったこと，腹部CTや血管撮影でも門脈構造が認められ，短絡血管自体の形態や太さからコイルによる塞栓術を選択した．

我々が検索した範囲では，今回のように先天性門脈体循環短絡症に対して，血管内コイル塞栓術により治療できた小児例の報告は少ない．まれな疾患ではあるが，治療のタイミングや治療方法を放射線科や小児外科など他科との綿密な検討を重ね慎重に門脈血流の評価を行うことで，合併症なく安全に閉塞が可能であると考えられた．

●文献

1) Morgan BG, Superina R : Congenital absence of the portal vein : Two cases and a proposed classification system for portasystemic vascular anomalies. J Pediatr Surg 1994 ; 29 : 1239-1241.

2) 井上勝昭, 酒本和也, 大竹治美, 他 : 大阪市における10年間のガラクトース血症要精密検査検体の診断結果について. 日本マス・スクリーニング学会誌 2012 ; 22 : 23-28.

3) Kim MJ, Ko JS, Seo JK, et al : Clinical features of congenital portosystemic shunt in children. Eur J Pediatr 2012 ; 171 : 395-400.

4) Uchino T, Matsuda I, Endo F : The long-term prognosis of congenital portosystemic venous shunt. J Pediatr 1999 ; 135 : 254-256.

5) Alonso-Gamarra E, Parron M, Perez A, et al : Clinical radiologic manifestations of congenital extrahepatic portosystemic shunts : A comprehensive review. Radio Graphics 2011 ; 31 : 707-722.

6) Kamimatsuse A, Onitake Y, Kamei N, et al : Surgical intervention for ductus venosus. Pediatr Surg Int 2010 ; 26 : 1025-1030.

7) Franchi-Abella S, Branchereau S, Lambert V, et al : Complications of congenital portosystemic shunts in children: therapeutic options and outcomes. J Pediatr Gastroenterol Nutr 2010 ; 51 : 322-330.

8) Schwartz YM, Berkowitz D, Lorber A, et al : Transvenous coil embolization of a patient ductus venousus in a 2-month-old child. Pediatrics 1999 ; 103 : 1045-1047.

9) Kim IO, Cheon JE, Kim WS, et al : Congenital intrahepatic portohepatic venous shunt : Treatment with coil embolization. Pediatr Radiol 2000 ; 30 : 336-338.

10) Pattynama P, Wils A, Linden E, et al : Embolization with the Amplatzer Vascular Plug in TIPS Patients. Cardiovasc Intervent Radiol 2007 ; 30 : 1218-1221.

11) Passalacqua M, Lie KT, Yarmohammadi H : Congenital extrahepatic portosystemic shunt (Abernethy malformation) treated endovascularly with vascular plug shunt closure. Pediatr Surg Int 2012 ; 28 : 79-83.

59

シトリン欠損による新生児肝内胆汁うっ滞症
(Neonatal Intrahepatic Cholestasis caused by Citrin Deficiency, NICCD) の診断と治療

大浦敏博[1), 2)]，坂本　修[2)]

1) 仙台市立病院小児科，2) 東北大学大学院医学系研究科小児病態学分野

はじめに

シトリンは肝ミトコンドリア膜に存在するアスパラギン酸・グルタミン酸キャリア (AGC) で，リンゴ酸・アスパラギン酸シャトルの一部を構成している．AGCは細胞質で生じたNADH還元当量のミトコンドリアへの輸送に必須である[1)]．シトリン欠損症ではAGCの機能低下による肝細胞質内のNADH蓄積が病態の根底にあると考えられている．糖負荷により細胞質NADHの過剰およびNAD$^+$の枯渇が生じるが，糖類を嫌う食癖はこれを避けるための自己防衛反応と考えられている．

シトリンをコードする遺伝子は*SLC25A13*であり，日本人患者では代表的な11個の変異で変異頻度の95%を占めている[2), 3)]．シトリン欠損症は東アジアから東南アジアで頻度が高く，少数ながら欧米からの報告もある．本邦での保因者頻度は1/65であり，理論上の有病率は1/17,000となる[2)]．

シトリン欠損症は年齢依存的に2つの病型を推移することが知られている (図1)．新生児から乳児の病型であるシトリン欠損による新生児肝内胆汁うっ滞症 (Neonatal Intrahepatic Cholestasis caused by Citrin Deficiency, NICCD) および成人期のシトルリン血症II型 (Citrullinemia type II, CTLN2) である[1)]．この間に「見かけ上健康」な適応・代償期がある．この時期の病態・病型をFTTDCD (Failure to thrive and dyslipidemia caused by citrin deficiency) と呼ぶ場合もある[4)]．本稿では小児期の病型であるNICCDの診断と治療について述べる．

1. NICCDの臨床像

新生児マススクリーニング (Newborn screening, NBS) 陽性を契機に生後1か月以内に精査受診となるもの (以下，NBS陽性群) と，NBSは正常で遷延性黄疸，体重増加不良などを契機に生後1~5か月の間に受診するもの (以下，NBS陰性群) に大別される[5)]．75例の検討ではNBS陽性群の割合は40%であった．

1) NBS陽性群

NBS陽性群は新生児期早期にすでに発症している群と考えられる．複数の血中アミノ酸 (シトルリン，チロシン，フェニルアラニン，メチオニン，スレオニンなど) や血中ガラクトースの上昇を呈することが多い．ガラクトースは初回検査が正常でも，遅れて上昇することが多いため，初回結果に関わらず再検時にはガラクトースも測定すべきである．ほとんどの症例で胆汁うっ滞，肝機能障害を認める．

2) NBS陰性群

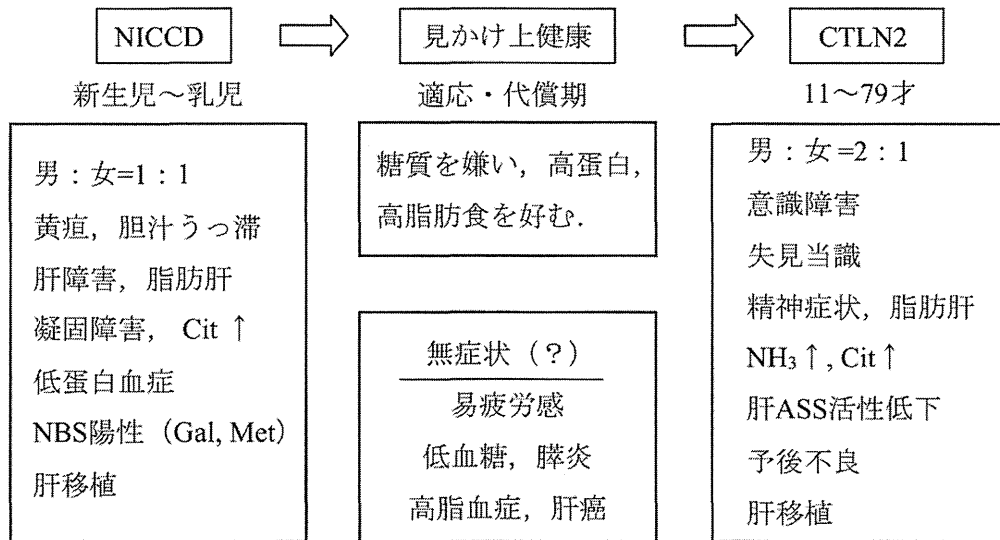この群のNBS結果は正常であり，NBS陽性群と比べてNICCDの発症が遅いと考えられる．多

受付日：平成27年3月2日
＜連絡先＞　大浦　敏博
〒982-8502　仙台市太白区あすと長町1丁目1-1
　　　　　仙台市立病院　小児科
　Tel：022-308-7111　Fax：022-308-7153
　E-mail：tohura@med.tohoku.ac.jp

NICCD	⇒	見かけ上健康	⇒	CTLN2
新生児〜乳児		適応・代償期		11〜79才

男：女=1：1
黄疸，胆汁うっ滞
肝障害，脂肪肝
凝固障害，Cit↑
低蛋白血症
NBS陽性（Gal, Met）
肝移植

糖質を嫌い，高蛋白，高脂肪食を好む．

無症状（？）
易疲労感
低血糖，膵炎
高脂血症，肝癌

男：女=2：1
意識障害
失見当識
精神症状，脂肪肝
NH_3↑, Cit↑
肝ASS活性低下
予後不良
肝移植

NICCD, neonatal intrahepatic cholestasis caused by citrin deficiency
CTLN2, Citrullinemia type 2
Cit, citrulline; Met, methionine; Gal, Galactose; NH_3, ammonia;
NBS, newborn screening; ASS, argininosuccinate synthetase

図1. シトリン欠損症における年齢依存性の二つの臨床病型

くの場合，遷延性黄疸，灰白色便を主訴に1か月以降に受診する．その他，肝腫大，出血傾向，低血糖，腹水，溶血性貧血などの報告がある．検査上は肝内胆汁うっ滞と肝機能障害が認められ，胆道閉鎖症や新生児肝炎との鑑別が重要となる[6]．NBS陽性群と異なりNBS陰性群では必ずしもアミノ酸やガラクトースの異常を呈さないことがある．

両群共通の臨床徴候として体重増加不良，凝固能低下，低蛋白血症，α-フェトプロテイン（AFP）高値，脂肪肝（腹部エコー，腹部CT，肝生検などで診断）が重要である．一部に肝障害が進行し，生体肝移植が必要であった症例が報告されている[7),8)]．

2. NICCDの診断

NICCDが疑われたら，血液・生化学一般検査に加え血中アミノ酸，ガラクトース，総胆汁酸，凝固能，AFPを測定する．また，エコー，CTなどで脂肪肝の有無を確認する．最終的な確定診断には遺伝子診断が必要である．表1にNICCDと鑑別すべき疾患をまとめた．

NBS陽性群ではその特徴的な症状・検査デー

タが揃えば，NICCDの診断は比較的容易である．タンデムマスの導入によりシトルリン高値で受診する場合も多いが，初診時には肝障害が目立たず，軽症のシトルリン血症1型との鑑別が難しい場合もある．必ず1〜2週間後にトランスアミナーゼや直接ビリルビンを再検すべきである．

NBS陰性群では，NBS陽性群に比して受診時期が数週遅いため血中アミノ酸やガラクトースの上昇の時期を逸していることがしばしばあり，NICCDの診断に至るまでに時間がかかる傾向がある．1歳までに肝不全が進行し，生体肝移植が施行された3症例はいずれもNBS陰性群で

表1. NICCDと鑑別すべき疾患

①新生児期に黄疸をきたす疾患
　1）胆道閉鎖症
　2）新生児肝炎

②新生児期に高ガラクトース血症をきたす疾患
　1）ガラクトース血症
　2）門脈体循環シャント

③シトルリン上昇をきたす疾患
　1）シトルリン血症Ⅰ型
　2）アルギニノコハク酸尿症

あり，肝移植時の病名はチロジン血症であった[7),8)]．早期治療介入により予後の改善が期待される疾患であり，原因不明の肝機能障害に遭遇した場合は，NICCDも鑑別に入れて検査を進めるべきである．

3. NICCDの治療

1) 乳糖制限・中鎖脂肪酸トリグリセリド (MCT) 添加乳

MCTは腸管内でミセルを形成せずに吸収され門脈血中に入り，カルニチンを必要とせずにミトコンドリア内に取り込まれ，β酸化を受けエネルギー源として利用される．その為，胆汁分泌障害のあるNICCDでも効率よくエネルギー産生が行える利点がある[9)]．必須脂肪酸強化MCTフォーミュラはMCTを含有し，乳糖も約半分に制限されているので最適である (表2-①)．

高ガラクトース血症を認める場合はガラクトースを除去する．乳糖除去ミルク100mlにMCTオイルを1〜2ml添加する (表2-②)．母乳は完全に止める必要はなく，特殊ミルクと併用しても良い (例えば母乳を1〜2回/日，他は特殊ミルク)．胆汁うっ滞，肝障害の改善が認められれば，特殊ミルクは漸減する．通常肝機能の正常化する1歳以降は，乳糖制限不要である．表2に使用可能な特殊ミルク一覧をしめす．明治721，明治110，森永ML-3は特殊ミルク事務局に申請

して入手する．

2) 脂溶性ビタミン経口補充

胆汁うっ滞が遷延する場合，脂溶性ビタミンの欠乏に注意する．投与量は新生児肝炎の治療に準ずる．

ビタミンA (チョコラ®A滴)
　100-500IU/kg/日
ビタミンD (アルファロール®内用液)
　0.01-0.1μg/kg/日
ビタミンE (ユベラ®顆粒)
　軽症 5-10mg/kg/日・中等症 20-50mg/kg/日・重症 50-100mg/kg/日
ビタミンK (ケイツー®シロップ)
　2mg/週 〜 5mg/日

3) 利胆剤

胆汁うっ滞が遷延する場合に併用．肝硬変，肝不全例には使用しない．

ウルソデオキシコール酸 (ウルソ®顆粒) (5-15mg/kg/日)

4) 新鮮凍結血漿, ビタミンK静注

著明な凝固能異常，出血傾向がある場合に併用する．

ケイツー®N静注　1回2〜5mg，1日1回静注

5) 肝移植

コントロールできない肝不全がある場合に考慮する．

表2. NICCDに用いられる特殊ミルク

①必須脂肪酸強化MCTフォーミュラ (明治721, 市販品あり)
　脂質の82%がMCTに置換され，炭水化物の組成は乳糖46%，
　可溶性多糖類54%となっている．
　P:F:C＝10:45:45, 71kcal/dl

②ガラクトース除去フォーミュラ (明治110, 市販品あり)
　P:F:C=12:37:51, 67.2kcal/dl
　上記100mLにMCTオイル2ml添加した場合　　P:F:C=9:51:40, 84kcal/dl
　　　　　　　　　　　　1ml添加した場合　　P:F:C＝10:45:45, 76kcal/dl

③蛋白質加水分解MCT乳 (森永ML-3)
　脂質の74%がMCTに置換され，可溶性多糖類と乳蛋白消化物から成る．
　P:F:C＝13:35:52, 70kcal/dl

(参考) 母乳　　P:F:C＝7:49:44, 65kcal/dl

P:F:C, 蛋白質, 脂質, 炭水化物の摂取カロリー比

4. 1歳以降のフォローアップについて

1) 低炭水化物食の励行

離乳期以降，糖質を嫌い，高蛋白・高脂肪食を好む食癖が認められる．この「特異な食癖」は，単なる「好き嫌い」ではなく，自己防衛反応であると考えられるため，それを矯正しないようにする．シトリン欠損患者の栄養調査によると摂取エネルギー比は蛋白：脂質：糖質＝15-25：40-50：30-40となることが知られている（日本人の食事の一般平均は15：25：60）[10]．糖質の過剰摂取を控える様に指導し，特に成人期のアルコール摂取は厳禁である．

2) 日常の受診および留意点

乳・幼児期は1～2ヵ月ごとに成長の確認（身長，体重），血算，一般生化学に加え血中アミノ酸，血糖，アンモニアをチェックする．学童期以降は4～6か月ごとに定期診察・検査を実施する．20歳以降は肝腫瘍の発生を念頭におき，数年に一回腹部エコーを行う．

体重減少，身長・体重の停滞，疲労感の増強，血中アンモニア値・シトルリン値・スレオニン／セリン値の上昇が認められた場合は注意が必要である．食事内容の再検討（糖質が過剰になっていないか検討する．

3) 特殊治療

易疲労感，体重増加不良，肝障害，高アンモニア血症，シトルリン血症などが改善しない場合は，以下の治療を考慮する．

①ピルビン酸ナトリウム（試薬）

細胞質内のNADHをNAD$^+$に変換することを目的にピルビン酸ナトリウム 0.1-0.3g/kg/日の投与が試みられ，体調の改善や食癖の変化が報告されている[11]．

②MCTオイル

この時期においてもMCTオイルが有効との報告がある[12]．

4) 末梢グルコース輸液

日常的に遭遇する低血糖，嘔吐下痢症などに用いられる末梢糖輸液製剤は低糖濃度であり差し支えない．この場合でも食事が可能になったら，早めに低炭水化物食を再開するようにする．

5) 禁忌：高カロリー輸液，濃グリセリン・果糖配合製剤（グリセオール®注など）

高濃度の糖液は肝細胞質のNADH/NAD$^+$比を上昇し，病態の悪化につながると考えられている[13],[14]．原因不明の肝不全，遷延性の凝固異常などでNICCDが鑑別にあがった場合には，（糖質による）高カロリー輸液を使用しない．

おわりに

シトリン欠損症は保因者頻度から患者は1/17,000人の頻度と計算されるが，実際に診断されるCTLN2は約1/100,000人であり，未発症例が存在すると考えられている．CTLN2発症の誘因は不明であるが，飲酒後に発症した症例の報告がある．またタンデムマス・スクリーニングでの発見率は1/80,000であり[15]，予想される頻度の1/4である．その為，多くの未診断（未発症）のシトリン欠損症が存在するものと考えられる．

シトリン欠損症は早期に診断することで，重篤なCTLN2の発症を予防することが可能である．原因不明の慢性肝障害，肝腫大に加え体重増加不良，易疲労感，低血糖などの症状を有する患児に遭遇した場合はシトリン欠損症も念頭に精密検査をすべきであろう．特異な食癖を確かめることが，診断上有用である．

文　献

1) Saheki T, Kobayashi K: Mitochondrial aspartate glutamate carrier (citrin) deficiency as the cause of adult-onset type II citrullinemia (CTLN2) and idiopathic neonatal hepatitis (NICCD). J Hum Genet. 47: 333-341, 2002

2) Tabata A, Sheng JS, Ushikai M, et al. Identification of 13 novel mutations including a retrotransposal insertion in SLC25A13 gene and frequency of 30 mutations found in patients with citrin deficiency. J Hum Genet. 53: 534-45, 2008

3) Kikuchi A, Arai-Ichinoi N, Sakamoto O, et al. Simple and rapid genetic testing for citrin deficiency by screening 11 prevalent mutations in SLC25A13. Mol Genet Metab. 105: 553-558, 2012

4) Song YZ, Zhang ZH, Lin WX, et al. SLC25A13 Gene Analysis in Citrin Deficiency: Sixteen Novel Mutations in East Asian Patients, and the Mutation Distribution in a Large Pediatric Cohort in China. PLoS One 8: e74544, 2013.

5) Ohura T, Kobayashi K, Tazawa Y, et al. Clinical pictures of 75 patients with neonatal intrahepatic cholestasis caused by citrin deficiency (NICCD). J Inherit Metab Dis. 30: 139-144, 2007

6) Tazawa Y, Kobayashi K, Abukawa D, et al. Clinical heterogeneity of neonatal intrahepatic cholestasis caused by citrin deficiency: case reports from 16 patients. Mol Genet Metab. 83: 213-219, 2004

7) Tamamori A, Okano Y, Ozaki H, et al. Neonatal intrahepatic cholestasis caused by citrin deficiency: severe hepatic dysfunction in an infant requiring liver transplantation. Eur J Pediatr. 161: 609-613, 2002

8) 中林啓記, 村上仁彦, 北澤恵美子ほか　乳児期に肝不全をきたし生体肝移植を施行したcitrin欠損症の1例　特殊ミルク情報 40: 30-35, 2004

9) Hayasaka K, Numakura C, Toyota K, et al. Treatment with lactose (galactose)-restricted and medium-chain triglyceride-supplemented formula for neonatal intrahepatic cholestasis caused by citrin deficiency. JIMD Rep. 2: 37-44, 2012

10) Saheki T, Kobayashi K, Terashi M, et al. Reduced carbohydrate intake in citrin-deficient subjects. J Inherit Metab Dis 31: 386-394, 2008

11) Mutoh K, Kurokawa K, Kobayashi K, et al. Treatment of a citrin-deficient patient at the early stage of adult-onset type II citrullinaemia with arginine and sodium pyruvate. J Inherit Metab Dis. 31(Suppl 2): S343-347, 2008

12) Hayasaka K, Numakura C, Toyota K, et al.: Medium-chain triglyceride supplementation under a low-carbohydrate formula is a promising therapy for adult-onset type II citrullinemia. Mol Genet Metab Reports 1: 42-50, 2014.

13) Yazaki M, Takei Y, Kobayashi K, et al. Risk of worsened encephalopathy after intravenous glycerol therapy in patients with adult-onset type II citrullinemia (CTLN2). Intern Med. 44: 188-195, 2005

14) Takahashi H, Kagawa T, Kobayashi K, et al. A case of adult-onset type II citrullinemia - deterioration of clinical course after infusion of hyperosmotic and high sugar solutions. Med Sci Monit. 12: CS13-15, 2006

15) 重松陽介. わが国におけるタンデムマス・スクリーニング試験研究の現状と課題. 日本先天代謝異常学会雑誌26(1) ; 31-34, 2010