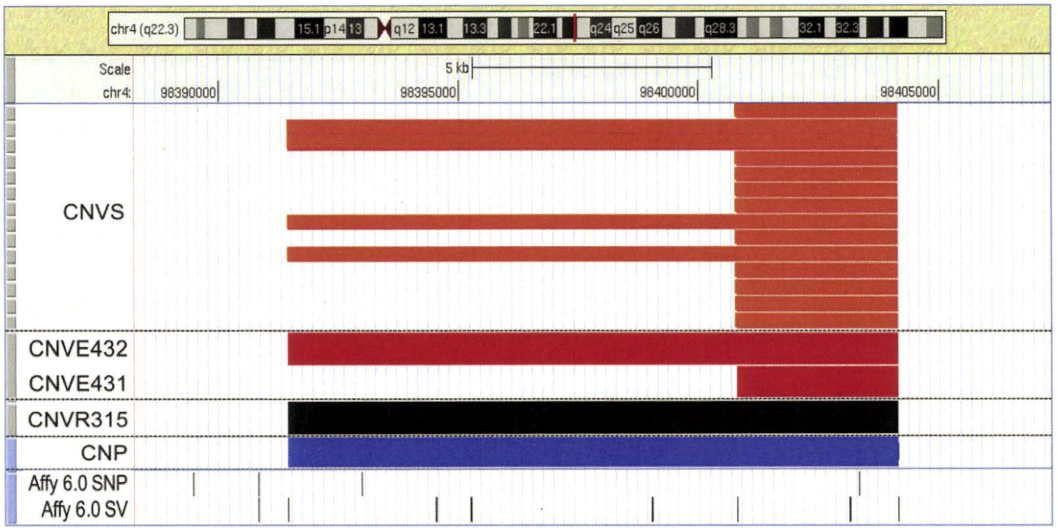

Figure 4. Map View of CNVRs Carrying CNVEs with Significant Haplotype Similarity
An example of a CNVR carrying CNVEs with significantly similar haplotype backgrounds is shown with the use of the UCSC Genome Browser. Other examples are presented in Figure S5. Thin bars in orange indicate the positions of CNVSs in individual CHMs. Thick bars in red, black, and blue represent the positions of CNVEs, CNVRs, and CNPs,[3] respectively. The bottom two lanes show the positions of SNP markers (Affy 6.0 SNP) and CNV markers (Affy 6.0 SV) in the Affymetrix SNP Array 6.0.

not explicitly stated, in previous reports.[3] McCarroll et al. demonstrated that most CNPs could be captured at a high linkage disequilibrium by nearby SNPs if the SNPs used were of sufficiently high density to allow estimation of the capture rate, despite the fact that some of the CNPs were clusters of CNVEs. These findings are most easily understood if haplotype-dependent recurrence of CNVEs is assumed. The possible dependence of CNVE occurrence on preexisting events is in contrast to SNPs, which can be regarded as the result of independent, random events.

The determination of CNV structure with the use of available arrays involves some uncertainty because of the extremely uneven distribution of markers, as noted previously.[3,19] Perhaps significant improvement in the detection of CNVs must await the availability of arrays carrying an unbiased distribution of markers. Recently, Conrad et al. reported an advanced CNV-typing array system that can efficiently detect even small CNVs.[27] With the use of this system, the detection of CNVs in existing materials should be improved; however, this system still suffers

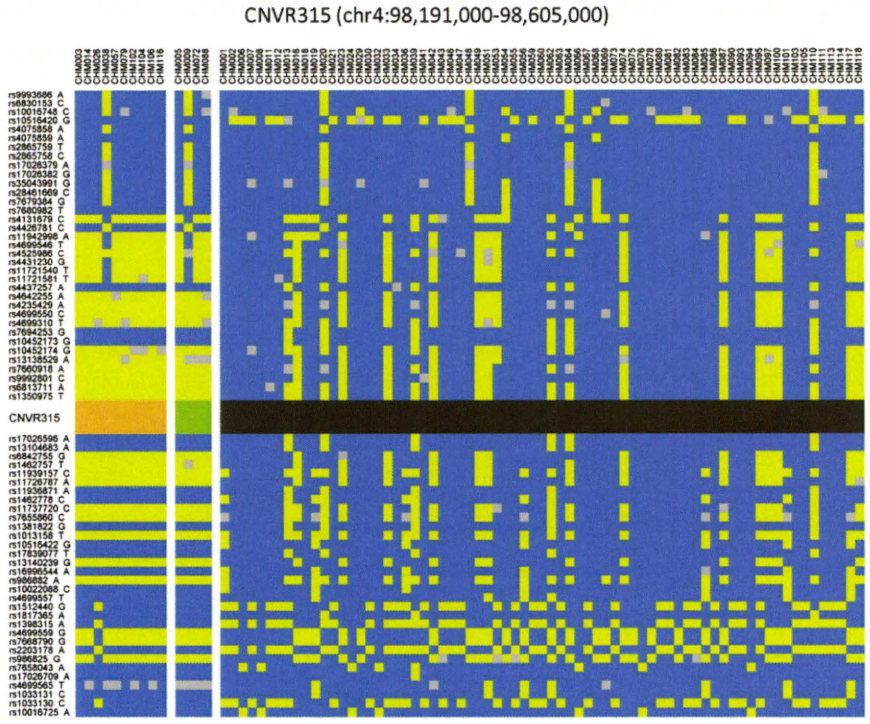

Figure 5. An Example of Haplotype Sharing between CNVEs
Haplotype profiles of CNVE samples (different CNVEs are color-coded by yellow or green in CNVR lines) and non-CNV samples (black in CNVR lines) for CNVR315 are shown. The major and minor SNP alleles are shown in blue and yellow, respectively, and SNPs with no genotype calls are shown in gray. See Figure S6 for the profiles of other CNVRs listed in Table 2.

from the fact that detecting CNVs in the Asian genome is highly inefficient (the number of CNVs detectable in Asians is approximately two-thirds that of individuals of European descent). This is because the initial experiments conducted to determine the markers to be loaded in the typing arrays were carried out with the use of European-descent and African samples, resulting in some population bias in the detection efficiency of the typing array.

Non-hybridization-based methods such as resequencing by new-generation sequencers are obviously among other future approaches. CHM samples provide an exceptional opportunity for effective whole-genome resequencing because CHMs display genome-wide homozygosity and require less sequencing redundancy. Furthermore, the reads can be aligned with greater confidence, unlike resequencing of diploid materials.

Supplemental Data

Supplemental Data include six figures and twelve tables and can be found with this article online at http://www.cell.com/AJHG.

Acknowledgments

We thank members of the Japan Association of Obstetricians & Gynecologists for their cooperation in collecting mole samples. We also thank Professor Yanagawa (Division of Biostatistics and Infectious Diseases, Kurume University School of Medicine, Kurume, Fukuoka) for help with the statistical evaluation of the haplotype preference of CNVEs. This work was supported by KAKENHI #17019051 (Grant-in-Aid for Scientific Research on Priority Areas "Applied Genomics"), KAKENHI #18710163 (Grant-in-Aid for Young Scientists [B]), and KAKENHI #20681020 (Grant-in-Aid for Young Scientists [A]) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, as well as by a grant from the Osaka Cancer Society.

Received: February 10, 2010
Revised: April 13, 2010
Accepted: May 7, 2010
Published online: May 27, 2010

Web Resources

The URLs for the data and software used herein are as follows:

Affymetrix: Genotyping Console software and annotation files, http://www.affymetrix.com/
Database of Genomic Variants, http://projects.tcag.ca/variation
dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/
Illumina: BeadStudio software and other requirement files, http://www.illumina.com/
R software, http://www.R-project.org
UCSC Genome Browser: genome annotation and SNP array marker information, http://genome.ucsc.edu/

Accession Numbers

The Gene Expression Omnibus (GEO) accession number for the array intensity data reported in this paper is GSE18701.

References

1. Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. (2004). Detection of large-scale variation in the human genome. Nat. Genet. *36*, 949–951.
2. Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., et al. (2004). Large-scale copy number polymorphism in the human genome. Science *305*, 525–528.
3. McCarroll, S.A., Kuruvilla, F.G., Korn, J.M., Cawley, S., Nemesh, J., Wysoker, A., Shapero, M.H., de Bakker, P.I., Maller, J.B., Kirby, A., et al. (2008). Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat. Genet. *40*, 1166–1174.
4. Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., et al. (2006). Global variation in copy number in the human genome. Nature *444*, 444–454.
5. Feuk, L., Marshall, C.R., Wintle, R.F., and Scherer, S.W. (2006). Structural variants: changing the landscape of chromosomes and design of disease studies. Hum. Mol. Genet. *15*(Spec No 1), R57–R66.
6. McCarroll, S.A. (2008). Extending genome-wide association studies to copy-number variation. Hum. Mol. Genet. *17*(R2), R135–R142.
7. Cook, E.H., Jr., and Scherer, S.W. (2008). Copy-number variations associated with neuropsychiatric conditions. Nature *455*, 919–923.
8. Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., et al; International HapMap Consortium. (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851–861.
9. Kidd, J.M., Cheng, Z., Graves, T., Fulton, B., Wilson, R.K., and Eichler, E.E. (2008). Haplotype sorting using human fosmid clone end-sequence pairs. Genome Res. *18*, 2016–2023.
10. Kukita, Y., Miyatake, K., Stokowski, R., Hinds, D., Higasa, K., Wake, N., Hirakawa, T., Kato, H., Matsuda, T., Pant, K., et al. (2005). Genome-wide definitive haplotypes determined using a collection of complete hydatidiform moles. Genome Res. *15*, 1511–1518.
11. Peiffer, D.A., Le, J.M., Steemers, F.J., Chang, W., Jenniges, T., Garcia, F., Haden, K., Li, J., Shaw, C.A., Belmont, J., et al. (2006). High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. Genome Res. *16*, 1136–1148.
12. Rozen, S., and Skaletsky, H.J. (2000). Primer3 on the WWW for general users and for biologist programmers. In Bioinformatics Methods and Protocols: Methods in Molecular Biology, S. Krawetz and S. Misener, eds. (Totowa, NJ: Humana Press), pp. 365–386.
13. Wang, T.L., Maierhofer, C., Speicher, M.R., Lengauer, C., Vogelstein, B., Kinzler, K.W., and Velculescu, V.E. (2002). Digital karyotyping. Proc. Natl. Acad. Sci. USA *99*, 16156–16161.
14. Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods *25*, 402–408.
15. Higasa, K., Kukita, Y., Kato, K., Wake, N., Tahira, T., and Hayashi, K. (2009). Evaluation of haplotype inference using definitive haplotype data obtained from complete

hydatidiform moles, and its significance for the analyses of positively selected regions. PLoS Genet. *5*, e1000468.
16. Zhang, J., Feuk, L., Duggan, G.E., Khaja, R., and Scherer, S.W. (2006). Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. Cytogenet. Genome Res. *115*, 205–214.
17. Bailey, J.A., Yavor, A.M., Massa, H.F., Trask, B.J., and Eichler, E.E. (2001). Segmental duplications: organization and impact within the current human genome project assembly. Genome Res. *11*, 1005–1017.
18. Sharp, A.J., Locke, D.P., McGrath, S.D., Cheng, Z., Bailey, J.A., Vallente, R.U., Pertz, L.M., Clark, R.A., Schwartz, S., Segraves, R., et al. (2005). Segmental duplications and copy-number variation in the human genome. Am. J. Hum. Genet. *77*, 78–88.
19. Cooper, G.M., Zerr, T., Kidd, J.M., Eichler, E.E., and Nickerson, D.A. (2008). Systematic assessment of copy number variant detection via genome-wide SNP genotyping. Nat. Genet. *40*, 1199–1203.
20. R Development Core Team. (2008). R: A language and environment for statistical computing (Vienna, Austria: R Foundation for Statistical Computing).
21. International HapMap Consortium. (2005). A haplotype map of the human genome. Nature *437*, 1299–1320.
22. Murdoch, S., Djuric, U., Mazhar, B., Seoud, M., Khan, R., Kuick, R., Bagga, R., Kircheisen, R., Ao, A., Ratti, B., et al. (2006). Mutations in NALP7 cause recurrent hydatidiform moles and reproductive wastage in humans. Nat. Genet. *38*, 300–302.
23. Slim, R., and Mehio, A. (2007). The genetics of hydatidiform moles: new lights on an ancient disease. Clin. Genet. *71*, 25–34.
24. Korn, J.M., Kuruvilla, F.G., McCarroll, S.A., Wysoker, A., Nemesh, J., Cawley, S., Hubbell, E., Veitch, J., Collins, P.J., Darvishi, K., et al. (2008). Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat. Genet. *40*, 1253–1260.
25. Perry, G.H., Ben-Dor, A., Tsalenko, A., Sampas, N., Rodriguez-Revenga, L., Tran, C.W., Scheffer, A., Steinfeld, I., Tsang, P., Yamada, N.A., et al. (2008). The fine-scale and complex architecture of human copy-number variation. Am. J. Hum. Genet. *82*, 685–695.
26. Pique-Regi, R., Ortega, A., and Asgharzadeh, S. (2009). Joint estimation of copy number variation and reference intensities on multiple DNA arrays using GADA. Bioinformatics *25*, 1223–1230.
27. Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., et al; Wellcome Trust Case Control Consortium. (2010). Origins and functional impact of copy number variation in the human genome. Nature *464*, 704–712.