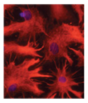
EXTENDED PDF FORMAT
SPONSORED BY



Travel Grants Available
Neuroscience 2014
Application Deadline: July 31st
>Apply Now
R&D systems
a biotechne brand
www.rndsystems.com

High-Resolution Genomic Analysis of Human Mitochondrial RNA Sequence Variation

Alan Hodgkinson *et al.*

Science **344**, 413 (2014);
DOI: 10.1126/science.1251110

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 4, 2014):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/content/344/6182/413.full.html

Supporting Online Material can be found at:
http://www.sciencemag.org/content/suppl/2014/04/23/344.6182.413.DC1.html

This article **cites 24 articles**, 10 of which can be accessed free:
http://www.sciencemag.org/content/344/6182/413.full.html#ref-list-1

This article appears in the following **subject collections:**
Genetics
http://www.sciencemag.org/cgi/collection/genetics

Downloaded from www.sciencemag.org on June 4, 2014

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2014 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

High-Resolution Genomic Analysis of Human Mitochondrial RNA Sequence Variation

Alan Hodgkinson,[1]* Youssef Idaghdour,[1,2]*† Elias Gbeha,[1] Jean-Christophe Grenier,[1] Elodie Hip-Ki,[1] Vanessa Bruat,[1] Jean-Philippe Goulet,[2] Thibault de Malliard,[1,2] Philip Awadalla[1,2]‡

Mutations in the mitochondrial genome are associated with multiple diseases and biological processes; however, little is known about the extent of sequence variation in the mitochondrial transcriptome. By ultra-deeply sequencing mitochondrial RNA (>6000×) from the whole blood of ~1000 individuals from the CARTaGENE project, we identified remarkable levels of sequence variation within and across individuals, as well as sites that show consistent patterns of posttranscriptional modification. Using a genome-wide association study, we find that posttranscriptional modification of functionally important sites in mitochondrial transfer RNAs (tRNAs) is under strong genetic control, largely driven by a missense mutation in *MRPP3* that explains ~22% of the variance. These results reveal a major nuclear genetic determinant of posttranscriptional modification in mitochondria and suggest that tRNA posttranscriptional modification may affect cellular energy production.

Mitochondria are vitally important for basic cellular function, and mutations in the compact genome of 16,569 base pairs (bp) have been associated with a large number of diseases in humans, as well as fundamental processes such as aging (*1*–*3*). Although regulation, transcription, and posttranscriptional processing of the human mitochondrial genome are subject to extensive research (*4*), very little is known about the nature, extent, and distribution of sequence variation in the mitochondrial transcriptome (*5*), particularly on a population level. Because it is ultimately translated variation that dictates downstream phenotypes, it is of great interest to investigate the extent and effects of such variation.

To deeply characterize sequence variation in the human mitochondrial transcriptome across a large number of individuals, we sequenced the whole-blood RNA of a random, predominantly French-Canadian population sample of 708 individuals from the CARTaGENE project (*6*) using the Illumina HiSeq platform. We then stringently aligned paired-end sequencing reads, focused on reads mapping to the mitochondria, and called variants using strict filters (*7*). On average, we achieved 6334× coverage (SD = 1139, range = 1456 to 9545) (fig. S1). Within each individual, we observe numerous sites that carry more than one allele (heteroplasmies), representing variation across mitochondrial transcriptomes (mtRNA). Between individuals, we observe high levels of mtRNA sequence variation, both in terms of the genomic locations of heteroplasmic sites and the proportions of alternative alleles carried at each site (Fig. 1).

On average, each individual carries 14.18 heteroplasmies identified in mtRNA across a total of 650 sites, higher than most previous estimates in human mitochondrial DNA (mtDNA) (*8*–*12*) (fig. S1, for biallelic site heteroplasmy annotation, see table S1). It is noteworthy that a number of positions are heteroplasmic within a high proportion of the 708 individuals (>25%); three of these sites consistently contain two alleles within each individual who exhibits heteroplasmy (positions C295T, G2129A, and G6691A in cDNA) and are thus potential candidates for RNA editing or modification. However, 13 sites are multiallelic (which here refers to sites containing three or more nucleotides) and often express all four nucleotides systematically within each individual. Sites that mismatch the genome—such that all four bases are observed—are indicative of reverse-transcription errors made at these sites during the process of creating and amplifying cDNA from RNA before sequencing and have been shown to overlap known posttranscriptionally modified sites (*13*–*16*).

Recently, 2 of the 13 multiallelic sites at positions 2617 (*RNR2*) and 13710 (*ND5*) were described as putative canonical RNA editing events (*5*). However, because we observe all four alleles at these sites across multiple individuals, it is likely that these changes actually represent posttranscriptional modifications. It is intriguing that the remaining 11 multiallelic sites all occur at the ninth position of transfer RNAs (tRNAs) (p9 sites), six of which are known to be posttranscriptionally methylated (*15*–*17*), which suggests that the presence of multiple alleles in our data represents posttranscriptional methylation at these sites.

We validated our results by sequencing both the whole mtDNA and two regions of mtRNA containing three p9 sites for five individuals using the IonTorrent platform to an average coverage of 9573× and 9633× for DNA and RNA, respectively, starting from stock DNA and RNA samples. In mtDNA, we find no evidence of alternative alleles above normal sequencing error rates (*7*) at the three p9 sites across all five individuals, whereas all four alleles are present in mtRNA on all occasions (table S2). Most important, the proportions of reads containing alternative alleles at these sites are significantly consistent across sequencing platforms (correlation coefficient $r^2 = 0.731$, $P = 4.89\text{e-}5$) (fig. S2). Because sequencing was performed after independent reverse transcription–polymerase chain reactions for each of the two platforms and the proportion of alternative alleles is systematic and repeatable across experiments, we hypothesize that this observed proportion represents the level of posttranscriptional methylation within each individual. Under this assumption, we observe between 0 and 74% methylation at p9 sites across the 708 individuals (fig. S3). We also validate positions 295, 2129, and 6691 as RNA modification events, because no heteroplasmies were found in DNA across the five individuals.

To determine whether methylation is occurring systematically for some mitochondrial organelles and not others, we examined whether alternative alleles at different p9 sites occur on the same read or paired sequencing reads more often than would be expected by chance. For the two positions close enough to be tested, considering the size distribution of our RNA sequencing (RNA-Seq) libraries, we see a significant enrichment of alternative alleles on the same background ($P < 0.001$) (fig. S2), which confirms that, when methylation takes place, it occurs consistently across mtRNA polycistronic molecules. By similar reasoning, we looked for correlations in the proportions of alternative alleles across p9 sites and observed significant correlations between 50 of the 55 pairwise comparisons ($P < 0.05$ after Bonferroni correction) (fig. S2).

Given the systematic, but highly variable, nature of methylation at p9 sites across individuals, we performed a genome-wide association study (GWAS) to test whether nuclear genetic variants are associated with the within-individual proportion of alternative counts at p9 sites. We genotyped all individuals using Illumina Omni2.5M single-nucleotide polymorphism (SNP) arrays and limited the analysis to SNPs with minor allele frequency (MAF) > 5%, in Hardy-Weinberg equilibrium ($P < 0.001$), and not missing in more than 1% of the individuals. We used all 11 p9 sites, calculating the combined alternative allele frequency for each individual, and considered a variety of models (*7*) (QQ plots shown in fig. S4).

The GWAS revealed a region on chromosome 14 that is associated with the within-individual proportion of alternative counts at genome-wide significance (Fig. 2, Table 1, and table S3). The strongest association occurs within the gene *MRPP3* at a missense mutation on the fourth

[1] CHU Sainte-Justine Research Centre, Department of Pediatrics, Faculty of Medicine, Université de Montreal, 3175 Chemin de la Côte-Sainte-Catherine, Montreal, Quebec H3T 1C5, Canada. [2] CARTaGENE, 3333 Queen Mary Road, Office 493, Montreal, Quebec H3V 1A2, Canada.

*These authors contributed equally to this work.

†Present address: Department of Biology, Division of Science and Mathematics, New York University Abu Dhabi, Post Office Box 129188, Abu Dhabi, United Arab Emirates.

‡Corresponding author. E-mail: philip.awadalla@umontreal.ca

exon (*rs11156878*, Asn → Ser, $P = 6.95\text{e-}34$). *MRPP3* codes for mitochondrial ribonuclease P protein 3 that forms part of mitochondrial RNase P enzyme complex, along with MRPP1 and MRPP2. This complex is involved in the cleavage and processing of mitochondrial transcripts for translation into proteins of the oxidative phosphorylation system (*18*). MRPP3 has been associated with the processing of the 5′ ends of tRNAs (*16*) and is suggested to contain a metallonuclease domain that would enable RNA hydrolysis (*18*). In vitro experiments have indicated that MRPP1 and MRPP2 may be involved in methylating mitochondrial tRNAs (*16*, *19*); however, MRPP3 has thus far not been implicated in this process. The GWAS did not detect significant associations within *MRPP1* or *MRPP2*, which suggests that genetic variation in *MRPP3* may modulate natural variation in the observed rate of posttranscriptional methylation in humans, either directly or via the processing activity of the mitochondrial RNase P enzyme.

Next, we performed a GWAS for each multiallelic site individually using the same association models as above and obtained the same variant (*rs11156878*) at genome-wide significance across eight p9 sites. We also identified four other genomic regions containing genome-wide significant markers that associate with the putative modification level at multiallelic sites (Table 1). Variation at positions 585 and 1610 show association with a missense variant within the gene *SLC25A26* (P values 3.83e-09 and 2.39e-17, respectively), which is a mitochondrial carrier protein, and for position 13710 (a non-p9 multiallelic site in *ND5*) with a missense variant within the gene *MTPAP* ($P = 2.03\text{e-}9$), which synthesizes the 3′ poly(A)$^+$ tail of mitochondrial transcripts. For position 2617 (a non-p9 multiallelic site in *RNR2*), the peak association falls upstream of *PPP1CB* ($P = 1.57\text{e-}11$); however, genome-wide significant associations also occur within *TRMT61B* (peak association has $P = 3.15\text{e-}11$), which is a methyltransferase that acts at position 58 of mitochondrial tRNAs and, therefore, may play a role in methylating *RNR2* at a functional position that has been linked to stability in the large ribosomal subunit (*5*).

To replicate the genetic associations in an independent data set, we performed RNA sequencing (average coverage 3029×) and genotyping using Omni 2.5M SNP arrays for an additional 287 individuals, and we find similar levels of variation in the mitochondrial genome (fig. S5). GWAS analyses replicate the peak SNP association in *MRPP3* for all p9 sites combined to a genome-wide significant level ($P = 2.56\text{e-}11$) (Table 1), which confirms the strength of this association. We also replicate the same peak SNP in *MRPP3* for three individual p9 sites, as well as an association linked to *SLC25A26*, but failed to replicate the associations with *MTPAP* and *PPP1CB* (Table 1 and table S4).

Variance explained by the missense variant *rs11156878* for all p9 sites combined is high ($r^2 = 22.03\%$) (Fig. 2), despite the relatively small sample size of our cohort. To fine-map the region most strongly associated with our phenotype, we sequenced the exomes of 96 individuals that were used in the original analysis and repeated the GWAS. We detected the strongest association signal with the same missense variant as before at experiment-wide significance level ($P = 3.85\text{e-}7$), which reflects the strength of the association.

It is noteworthy that the frequency distribution of *rs11156878* is not consistent across worldwide populations; specifically, in 1000 Genomes data (*20*), the minor allele frequency shows variation across continents (14% in the Americas, 17% in Europe, 5% in Asian populations, and 4% in African populations). The specific effect of this variant is not known; however, given that the level of methylation at p9 sites has been linked to levels of protein translation and mitochondrial respiration (*16*), variability at this site across worldwide populations may have significant implications on region-specific health and disease.

Finally, given the multiallelic variability across p9 sites, we performed a multiple regression between the putative level of methylation at p9 sites and the basal metabolic rate (BMR) of individuals measured using a body composition analyzer (*7*). BMR represents the total energy expended by the body to maintain normal functions at rest, such as respiration and circulation. For all p9 sites combined, there is no significant association between BMR and the proportion of alternative alleles at p9 sites ($P = 0.150$). However, for two individual p9 sites we find suggestive evidence for a positive association (position 7526, *TRND*, $P = 0.0034$ and position 14734, *TRNE*, $P = 0.0084$), one of which is replicated in our independent data set of 287 individuals (position 7526, *TRND*, $P = 0.0192$). Resolving the significance of any association implicating RNA sequence variation in the basis of organismal phenotypes requires systematic investigations; however, these observations, together with those that highlight the importance of posttranscriptional methylation for the correct folding of tRNA molecules (*21*, *22*), suggest that

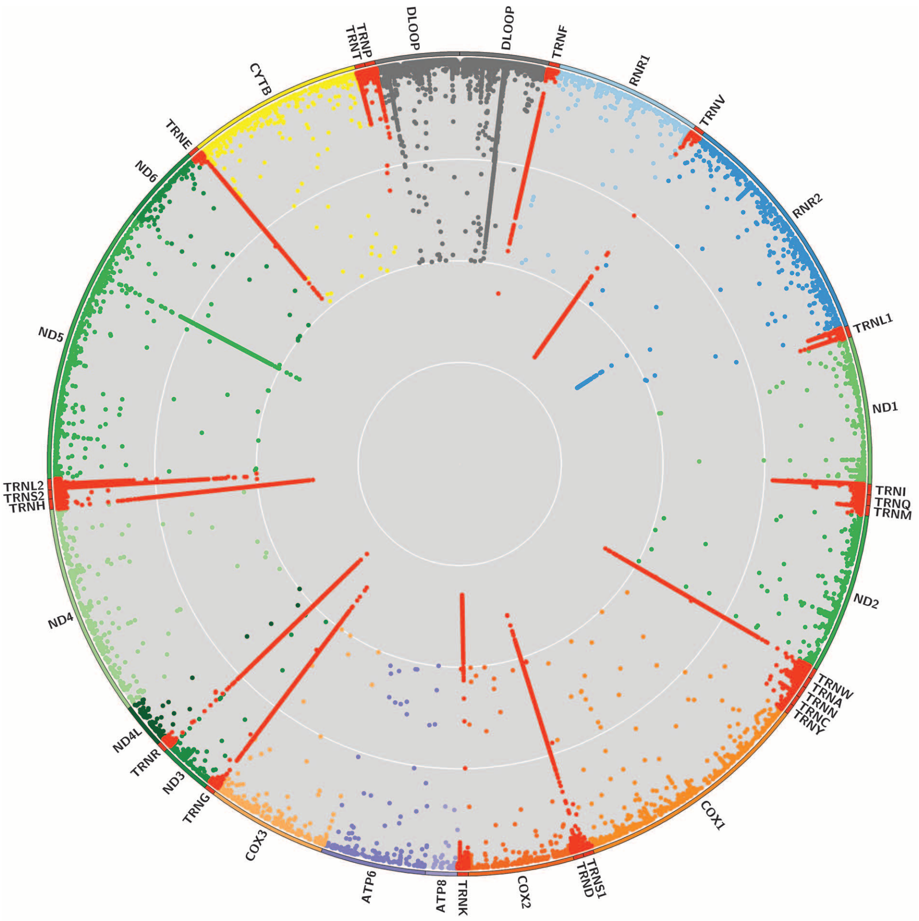


Fig. 1. Polymorphism information content (PIC) for each site in the mitochondrial transcriptome across 708 individuals. PIC is calculated as $1-\sum_{i=1}^{n} p_i^2$, where p_i is the proportion of sequencing reads containing each nucleotide within an individual and n is the number of different alleles at each site. Each data point therefore represents the PIC value for a single site using allele frequency information from sequencing reads within a single individual. White circular lines represent PIC scores of 0.25, 0.5, and 0.75, from outside to in. Sites with PIC < 0.005 are not shown. For sites containing two alleles, the maximum PIC score is 0.5.

investigating the role posttranscriptional tRNA sequence variation in pathologic contexts is likely to be revealing.

Mitochondrial tRNAs are vitally important to mitochondrial function; to date, ~250 tRNA mutations have been associated with disease (*23*). Similarly, posttranscriptional events in RNA appear to be common (*5*, *24*) and may be linked to biological function and disease (*25*). It was previously shown that modification events at p9 sites could be detected in vitro using cells lines (*15*, *16*); however, here we characterize the nature and extent of the between-individual variation in vivo across a large number of individuals at mitochondrial genome-wide scale. Furthermore, we uncovered a previously underappreciated role of *MRPP3*, likely through missense variant *rs11156878*, in modulating the rate of posttranscriptional modification, either directly or through the processing activities of mitochondrial RNase P. These results warrant further characterization of the mechanism of action of MRPP3 and its potential role in shaping metabolic-related processes in humans. More broadly, our study demonstrates the potential of GWAS to identify genes involved in basic cellular pathways and highlights the use of RNA sequence variation to complement DNA studies to assess variation and its potential consequences at the population level.

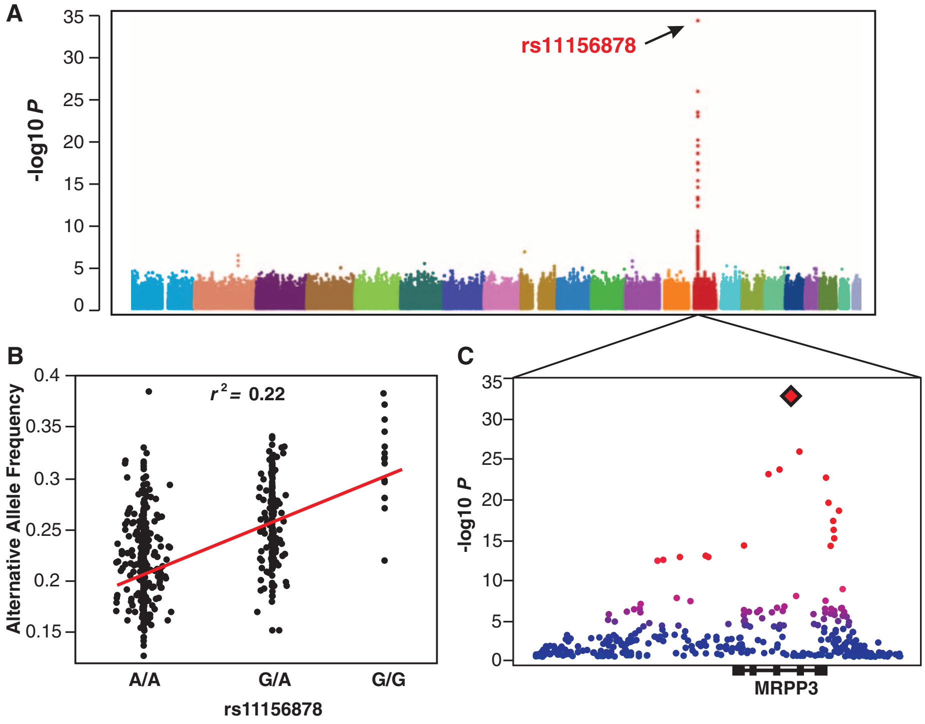


Fig. 2. GWAS results for all p9 sites combined. The $-\log_{10} P$ values are shown for all sites across the nuclear autosomal genome with each chromosome in a difference color. (**A**) The full model [model 3, see (*7*)] accounts for age, gender, site of collection, global genetic ethnicity, and complete blood counts. (**B**) The correlation between minor allele count and combined alternative allele frequency is shown for *rs11156878*. (**C**) A magnified section of the full GWAS plot containing the gene *MRPP3*.

Table 1. GWAS results for the full linear model at multiallelic sites. The peak SNP is shown for each region that contains genome-wide significant markers in each test during the discovery phase. The full linear model corrects for age, gender, site of collection, global genetic ethnicity, and complete blood counts. ALL p9 refers to resampled data from all multiallelic p9 sites combined.

Trait	Peak SNP	Position	MAF	Gene	Annotation	Change	*P value*	
							Discovery	Replication
ALL p9	*rs11156878*	14:35735967	0.1939	*MRPP3*	Missense	Asn→Ser	6.95×10^{-34}	2.56×10^{-11}*
1610	*rs11156878*	14:35735967	0.1796	*MRPP3*	Missense	Asn→Ser	7.25×10^{-10}	3.93×10^{-04}
5520	*rs11156878*	14:35735967	0.1807	*MRPP3*	Missense	Asn→Ser	4.89×10^{-16}	5.80×10^{-05}
8303	*rs11156878*	14:35735967	0.1839	*MRPP3*	Missense	Asn→Ser	7.44×10^{-12}	9.14×10^{-03}
9999	*rs11156878*	14:35735967	0.1805	*MRPP3*	Missense	Asn→Ser	2.24×10^{-33}	8.36×10^{-10}*
10413	*rs11156878*	14:35735967	0.1800	*MRPP3*	Missense	Asn→Ser	1.16×10^{-26}	9.19×10^{-06}
12146	*rs11156878*	14:35735967	0.1808	*MRPP3*	Missense	Asn→Ser	1.18×10^{-31}	5.25×10^{-09}*
12274	*rs11156878*	14:35735967	0.1802	*MRPP3*	Missense	Asn→Ser	3.77×10^{-21}	1.08×10^{-06}*
14734	*rs11156878*	14:35735967	0.1796	*MRPP3*	Missense	Asn→Ser	5.73×10^{-11}	1.25×10^{-03}
585	*rs13874*	3:66419956	0.4444	*SLC25A26*	Missense	Thr→Met	3.83×10^{-09}	1.32×10^{-05}
1610	*rs13874*	3:66419956	0.4438	*SLC25A26*	Missense	Thr→Met	2.39×10^{-17}	6.98×10^{-08}*
2617	*rs12714241*	2:28969413	0.3554	*NA*	Intergenic	T→C	1.57×10^{-11}	0.36870
13710	*rs1047991*	10:30629226	0.2512	*MTPAP*	Missense	Arg→Cys	2.03×10^{-09}	0.08675

*The site is also the peak SNP in the replication phase.

References and Notes

1. J. M. Ross *et al.*, *Nature* **501**, 412–415 (2013).
2. R. W. Taylor, D. M. Turnbull, *Nat. Rev. Genet.* **6**, 389–402 (2005).
3. D. C. Wallace, *Science* **283**, 1482–1488 (1999).
4. O. Rackham, T. R. Mercer, A. Filipovska, *RNA* **3**, 675–695 (2012).
5. D. Bar-Yaacov *et al.*, *Genome Res.* **23**, 1789–1796 (2013).
6. P. Awadalla *et al.*, *Int. J. Epidemiol.* **42**, 1285–1299 (2013).
7. Materials and methods are available as supplementary materials on *Science* Online.
8. H. Goto *et al.*, *Genome Biol.* **12**, R59 (2011).
9. Y. He *et al.*, *Nature* **464**, 610–614 (2010).
10. M. Li *et al.*, *Am. J. Hum. Genet.* **87**, 237–249 (2010).
11. B. A. Payne *et al.*, *Hum. Mol. Genet.* **22**, 384–390 (2013).
12. D. C. Samuels *et al.*, *PLOS Genet.* **9**, e1003929 (2013).
13. H. A. Ebhardt *et al.*, *Nucleic Acids Res.* **37**, 2461–2470 (2009).
14. S. Findeiss, D. Langenberger, P. F. Stadler, S. Hoffmann, *Biol. Chem.* **392**, 305–313 (2011).
15. T. R. Mercer *et al.*, *Cell* **146**, 645–658 (2011).
16. M. I. G. Sanchez *et al.*, *Cell Cycle* **10**, 2904–2916 (2011).
17. M. A. Machnicka *et al.*, *Nucleic Acids Res.* **41**, (D1), D262–D267 (2013).
18. J. Holzmann *et al.*, *Cell* **135**, 462–474 (2008).
19. E. Vilardo *et al.*, *Nucleic Acids Res.* **40**, 11583–11593 (2012).
20. G. R. Abecasis *et al.*, *Nature* **491**, 56–65 (2012).
21. M. Helm *et al.*, *Nucleic Acids Res.* **26**, 1636–1643 (1998).
22. M. Helm, R. Giegé, C. Florentz, *Biochemistry* **38**, 13338–13346 (1999).
23. E. Ruiz-Pesini *et al.*, *Nucleic Acids Res.* **35**, (Database), D823–D828 (2007).
24. E. Y. Levanon *et al.*, *Nat. Biotechnol.* **22**, 1001–1005 (2004).
25. N. Paz *et al.*, *Genome Res.* **17**, 1586–1595 (2007).

Acknowledgments: A.H., Y.I., and P.A. designed the study. P.A. provided samples; E.G. and E.H.-K. processed samples for sequencing and genotyping; J.-C.G. and V.B. aligned sequencing data. Y.I., A.H., T.d.M., and J.-C.G. performed quality control on sequencing data, and Y.I. and J.-P.G. performed quality control on genotyping data. J.-P.G. compiled phenotype data from the CARTaGENE project. A.H. and Y.I. performed the analysis. A.H., Y.I., and P.A. wrote the manuscript. We thank E. Shoubridge, I. Thiffault, G. Gibson, J. Novembre, and E. Stone for comments on the study; C. Privé for help with study design on the Ion Torrent platform; and the CARTaGENE team for assistance with data collection. We acknowledge financial support from Fonds de la recherche en santé du Québec (FRSQ), Genome Quebec, Réseau de médecine génétique appliquée (RMGA), Fonds québécois de la recherche sur la nature et les technologies (FQRNT), Canadian Foundation of Innovation, and the Canadian Partnership Against Cancer. A.H. is an FRSQ Research Fellow and Y.I. is a Banting Postdoctoral Fellow. Requests for data published here should be submitted to the CARTaGENE program at access@cartagene.qc.ca, or the corresponding author, citing this study. The authors declare no conflict of interests.

Supplementary Materials

www.sciencemag.org/content/344/6182/413/suppl/DC1
Materials and Methods
Figs. S1 to S7
Tables S1 to S8
References

21 January 2014; accepted 26 March 2014
10.1126/science.1251110