A Māori specific *RFC1* pathogenic repeat configuration in CANVAS, likely due to a founder allele

Running head: Māori specific *RFC1* pathogenic repeat configuration

Authors: Sarah J. Beecroft[1,2]*, Andrea Cortese[3]*, Roisin Sullivan[3+], Wayne Yau[3+], Zoe Dyer[4], Teddy Y. Wu[5], Eoin Mulroy[4], Luciana Pelosi[4], Miriam Rodrigues[4], Rachael Taylor[6], Stuart Mossman[7], Ruth Leadbetter[7], James Cleland[8], Tim Anderson[5], Gianina Ravenscroft[1,2], Nigel G. Laing[1,2], Henry Houlden[3], Mary M. Reilly[3], Richard H. Roxburgh[4,6]

*Joint first

+Joint second

Affiliations:

1 Neurogenetic Diseases Group, Centre for Medical Research, QEII Medical Centre, University of Western Australia, Nedlands, WA 6009, Australia

2 Harry Perkins Institute of Medical Research, QEII Medical Centre, Nedlands, WA 6009, Australia

3 Department of Neuromuscular Disease, UCL Queen Square Institute of Neurology and The National Hospital for Neurology and Neurosurgery, London, UK

4 Neurology Department, Auckland City Hospital, Private Bag 92024, Auckland, New Zealand

5 Department of Neurology, Christchurch Hospital, Christchurch, New Zealand

6 Centre for Brain Research Neurogenetics Research Clinic, University of Auckland, Auckland, New Zealand

7 Neurology Department, Wellington Hospital, Wellington, New Zealand

8 Neurology Department, Tauranga Hospital, Tauranga, New Zealand

Corresponding Authors:

Sarah J. Beecroft
Harry Perkins Institute of Medical Research, QEII Medical Centre

Nedlands, WA 6009, Australia
Email: **sarah.beecroft@uwa.edu.au**

Richard H. Roxburgh
Neurology Department, Auckland City Hospital
Private Bag 92024, Auckland, New Zealand
Email: **RichardR@adhb.govt.nz**

Abstract

Cerebellar ataxia with neuropathy and bilateral vestibular areflexia syndrome (CANVAS) is a recently recognised neurodegenerative disease with onset in mid- to late adulthood. The genetic basis for a large proportion of Caucasian patients was recently shown to be the biallelic expansion of a pentanucleotide $(AAGGG)_n$ repeat in *RFC1*. Here, we describe the first instance of CANVAS genetic testing in New Zealand Māori and Cook Island Māori individuals. We show a novel, possibly population-specific CANVAS configuration $(AAAGG)_{10\text{-}25}(AAGGG)_{exp}$, that was the cause of CANVAS in all patients. There were no apparent phenotypic differences compared with European CANVAS patients. Presence of a common disease haplotype among this cohort suggests this novel repeat expansion configuration is a founder effect in this population, which may indicate that CANVAS will be especially prevalent in this group. Haplotype dating estimated the most recent common ancestor at approximately 1430CE. We also show the same core haplotype as previously described, supporting a single origin of the CANVAS mutation.

Keywords:

CANVAS; repeat expansion; RFC1; founder effect; Māori

Abbreviations:

BAM file: Binary Aligned/Mapped file; CANVAS: Cerebellar ataxia with neuropathy and bilateral vestibular areflexia syndrome; GATK4: Genome analysis tool kit version 4; MCRA:

Most recent common ancestor; NGS: next-generation sequencing; WES: whole exome sequencing; WGS: whole genome sequencing

Introduction

The combination of cerebellar ataxia, neuropathy, and bilateral vestibular areflexia was recently recognised as a distinct syndrome (CANVAS) (Szmulewicz *et al.*, 2011). This slowly progressive neurodegenerative disease usually presents in mid to late adulthood (>30 years) (Szmulewicz *et al.*, 2011). Additional features include chronic cough, autonomic dysfunction (Taylor *et al.*, 2014; Cortese *et al.*, 2019) and thinning of the peripheral nerves (Pelosi *et al.*, 2018). Striking neuropathological features include atrophy of the Purkinje cells; vestibular, geniculate and trigeminal ganglia; and dorsal root ganglia and posterior columns (Szmulewicz *et al.*, 2014). Cortese *et al*. (2019) showed that a recessive pentanucleotide repeat expansion is responsible for the vast majority of Caucasian cases and this has been confirmed in a second cohort from Australia (Rafehi *et al.*, 2019). The expansion occurs in the poly(A) tail of an AluSx3 element in intron two of *RFC1*. At this locus, the reference allele is $(AAAAG)_{11}$. Other benign configurations include $(AAAAG)_{exp}$ and $(AAAGG)_{exp}$. The pathogenic CANVAS allele is estimated to be 400-2,000 repeated units of AAGGG. The estimated carrier frequency of the pathogenic CANVAS allele is 0.7% in Caucasians, which would make CANVAS one of the most common hereditary forms of late-onset ataxia (Cortese *et al.*, 2019). However, this finding requires validation in other populations. We describe the first reported genetic characterisation of CANVAS in New Zealand Māori and Cook Island Māori individuals, who comprise a significant part of the New Zealand/Cook Island population. CANVAS has been seen in this group previously (Taylor *et al.*, 2014). We show that these patients have a different conformation of the pathogenic pentanucleotide repeat.

Materials and methods

Cohort

This study was approved by the University of Western Australia Human Research Ethics Committee and the New Zealand Health and Disability Ethics Committee. DNA was obtained for 15 individuals; 13 were affected, and 2 were unaffected family members. All individuals gave informed consent for DNA collection and analysis, except M2 III:2. M2 III:2 passed away prior to the study, and therefore informed consent was obtained from the patient's family to use a stored DNA sample. Affected individuals were recruited from neurology clinics in New Zealand. There were two multiplex families and five singleton patients (Fig. 1 and Table 1). Twelve individuals were Māori (M), and three were Cook Island Māori (CI). Five individuals (M3 I:1, M4 I:1, M5 I:1 M6 I:1 M7 I:1) have been previously described clinically (Taylor *et al.*, 2014).

Clinical Data Collection

Apart from M2 III:2, all affected individuals were clinically assessed at study recruitment. Follow-up information was taken from clinical documentation, phone interviews and follow-up examination of key affected individuals (n=13). Contact was made with seven affected individuals. Four patients could not be contacted, and two affected individuals (M2 III:5 and M3 I:1) had died.

Genetic testing

To determine if affected individuals were homozygous for the CANVAS pathogenic repeat expansion, we followed the protocols described in Cortese *et al.*(2019). Briefly, we performed standard PCR with primers flanking the CANVAS locus. The absence of any product suggests a homozygous expansion that is too large to be amplified by standard PCR. One band indicates that at least one allele is within the amplifiable range of normal PCR, and

therefore the individual is not homozygous for a CANVAS pathogenic expansion. The expected product size is 355bp, which corresponds to the reference $(AAAAG)_{11}$. A larger band indicates an intermediate sized expansion, larger than the reference allele but smaller than a full pathogenic expansion. For individuals where absence of a product upon flanking PCR suggested an expansion, we then performed repeat-primed PCR (RP-PCR), using primers specific for the AAAAG (reference allele), AAAGG (benign variant), and AAGGG (pathogenic) configurations. Due to individual M2 V:1's unusually early clinical presentation, the Fulgent ataxia repeat expansion panel (*ATN1, ATXN1, ATXN10, ATXN2, ATXN3, ATXN7, ATXN8, ATXN8OS, BEAN1, CACNA1A, FMR1, FXN, NOP56, PPP2R2B, TBP*), and a broad neurogenetic gene panel (Beecroft *et al.*, 2020) (genes listed in Supplementary Table 1) was also performed to search for a possible other genetic condition explaining her early disease onset. In addition to these we performed functional studies to exclude Wilson's disease (copper studies), ataxia telangiectasia (alpha fetoprotein and ATM protein/kinase activity), vanishing white matter disease (transferrin isoforms) and white-cell enzyme testing.

Next-generation sequencing

Illumina whole genome sequencing (WGS) was performed on individuals CI1 II:1, C1 II:3, M2 III:4, and M2 V:1 via the Australian Genome Research Facility (Melbourne), following GATK4 best-practices (Poplin *et al.*, 2017). We compared this data against the selected markers from the haplotyping analysis by Cortese *et al.* (2019) (spanning chr4: 38157510- 40712481, hg19). Illumina whole exome sequencing (WES) was performed on individuals M3 I:1, M4 I:1, M5 I:1, M6 I:1 and M7 I:1 via the Australian Genome Research Facility (Melbourne), following GATK4 best-practices (Poplin *et al.*, 2017). Using Linkdatagen (Bahlo and Bromhead, 2009), informative HapMap2 markers were extracted from the combined exome and genome sequencing and prepared for analysis with Merlin (Abecasis

et al., 2002). Markers were excluded if they were covered to a read depth of <20-fold, or not sequenced in ≥50% of samples. We used Merlin to generate the most likely haplotypes. The length of the shared homozygous haplotype in centimorgans was calculated from the Merlin output. The haplotype lengths were used to calculate the most recent common ancestor assuming a 'correlated' genotype (95% confidence interval) using the Genetic Mutation Age Estimator tool (https://shiny.wehi.edu.au/rafehi.h/mutation-dating/) (Gandolfo *et al.*, 2014), employed in a recent haplotype analysis of CANVAS patients (Rafehi *et al.*, 2019).

Data availability

Anonymised data is available from the corresponding author.

Results

Cohort

The summarised clinical features of the affected individuals are presented in Table 2, which are compared with a cohort of 16 New Zealand European CANVAS patients (Cortese *et al.* 2019, in press, Brain), and with the cohort published by Cortese *et al*. (2019). Clinical features were similar across all three patient groups. Detailed clinical information for our cohort is shown in Supplementary Table 1.

The mean age of symptom onset was 55 years (excluding individual M2 V:1). The presenting neurological symptom was unsteadiness in all but one affected individual (M2 III:4), where dysarthria predated unsteadiness by two years. Nine affected individuals described persistent cough, which predated neurological symptoms in three affected individuals. A third of patients complained of vestibular. In their initial neurological examination, seven affected individuals showed nystagmus. Ten affected individuals required a walker during their illness. The mean time between symptom onset and the use of a walking

frame was 7 years (range 1-15 years). Two affected individuals died following study recruitment. M2 III:5 experienced sudden cardiac death, but the cause of death for M3 I:1 was unknown.

Investigations

Eight of thirteen patients had electrodiagnostic studies. In six, motor nerve conduction studies were normal, and in two there was mild reduction in motor conduction velocity or amplitude. Nine affected individuals had cerebellar atrophy on initial MRI brain scan. M2 III:4 had a normal brain MRI three years after symptom onset, but subsequent CT head scans showed cerebellar atrophy. Video head impulse tests were consistent with bilateral vestibulopathy in all eleven tested patients including those with a normal clinical head impulse test at initial neurological assessment. Individual M2 III:5 (*RFC1* pathogenic repeat positive) was asymptomatic at age 59 years, excepting chronic cough with syncope, and painful extremities. He showed mild bilateral vestibular impairment, but no further examinations were performed. He died from sudden cardiac death before he could be re-examined. M3 I:1, M4 I:1, M5 I:1 had formal autonomic nervous system testing. Abnormalities were inconsistently in the sympathetic and parasympathetic systems in different patients. Ultrasound examination showed small nerve cross-sectional areas in M4 I:1, M6 I:1, M7 I:1, but not M2 V:1.

Individual M2 V:1

Of particular interest was individual M2 V:1. She presented at age 6 with tremor, gait unsteadiness, and learning difficulties. Tremor was reportedly present since infancy. Otherwise, motor milestones were normal. She had prominent upper limb intention tremor and impaired tandem gait. Communication was significantly impaired by dysarthria. Cognition was difficult to assess formally. She had no nystagmus or sensory deficit, but upper neurone signs

with increased tone and extensor plantars. Brain MRI at age 6 showed increased signal in the cerebellar peduncles and dorsal brainstem (Supplementary Fig. 2A). A limited follow up scan at the age of 15 showed particular prominence of the horizontal fissure, as is seen in crus 1 cerebellar atrophy of CANVAS, together with enlarged supratentorial CSF spaces consistent with a degree of more widespread atrophy (Supplementary Fig. 2B). The scan was limited due to movement so the white matter changes seen previously were unable to be further interrogated. She was wheelchair bound at age 8. At age 20, she additionally showed profound limb ataxia as measured using the finger-nose-finger, finger chase and heel shin tests of the scale for the assessment and rating of ataxia (SARA); her SARA score was 31. Bilateral vestibular failure was demonstrated on video head impulse test, with a gain of 0.46 (left) and 0.49 (right). Nerve conduction studies (age 20) were atypical for CANVAS showing mild, uniform slowing of conduction velocities but normal sensory and motor nerve amplitudes. Extensive investigation for other causes of ataxia (including Friedreich's ataxia) and cognitive impairment revealed no alternative cause. The neurogenetic gene panel and Fulgent ataxia panel revealed no likely pathogenic variants, and all functional testing was normal.

Flanking PCR

There was no PCR product in 13/15 individuals (Table 1), indicating they harboured an expansion on both alleles. Of the two remaining individuals, M2 IV:1 showed a ~355bp band, indicating one normal-sized allele. CI1 II:2 showed a larger band, indicating an intermediate sized allele that was expanded. All affected individuals had no PCR amplifiable product, while the two unaffected individuals did show a PCR product. Thus, the expanded allele segregated with disease.

Repeat-primed PCR

All thirteen individuals without a band on flanking PCR were also negative on repeat-primed PCR for the reference pentanucleotide sequence (AAAAG) at the CANVAS locus. These individuals shoed the pathogenic AAGGG repeat, as seen in the European CANVAS population (Fig. 2B). However, they were also found to have a novel configuration at the locus. The pathogenic AAGGG expansion was preceded by a stretch of AAAGG benign variant repeats (Fig. 2B), which varied in number. The individuals had the conformation $(AAAGG)_{10\text{-}25}(AAGGG)_{exp}$. The carriers M2 IV:1 and CI1 II:2 each harboured one allele of this configuration. On the other allele, M2 IV:1 had the reference sequence $(AAAAG)_{11}$, while CI1 II:2 had an intermediate sized AAAGG expansion.

Next-generation sequencing

Visual interrogation of soft-clipped reads in the BAM files of the WGS for the three individuals CI1 II:2, M2 III:2, and M2 IV:1 showed a small number of repeats of the benign variant allele $(AAAGG)_{4\text{-}6}$ at the distal end of the *RFC1* pathogenic expansion in the reads that covered this region (Fig. 2C). The coverage of this region in CI1 II:1 was too low to reliably detect this pattern. The disease-associated repeat expansion in the Maori and Cook Island Maori is thus a hybrid allele of the pathogenic AAGGG repeat embedded in benign variant AAAGG repeats (Fig. 2A).

Haplotype analysis

Cortese *et al.* (2019) identified a shared haplotype in their patients, encompassing a 47.9kb core region that was identical in all but one patient (chr4:39318706-39366590, hg19). Our four patients with WGS data shared this core haplotype, plus an additional region extending from chr4:39122697-39366590 (0.24Mb total shared) (Supplementary Table 2). We then combined the WES and WGS data to impute the most likely haplotypes for our patients

(Supplementary Table 3). This showed an extended shared haplotype in our cohort, spanning chr4:36317970-44295839 (7.98Mb; Supplementary Table 3). Based on the length of the shared haplotype in centimorgans, the estimated most recent common ancestor was 25.6 generations ago (95% CI 11.1-60.7). Assuming that one generation is 25 years, the most recent common ancestor (MRCA) was 650 years ago (95% CI 275-1525; dating to 1369 CE, 95% CI 494-1744 CE). Assuming instead that one generation is 20 years, this would place the most recent common ancestor 520 years ago (95% CI 220-1200 years; dating to 1499 CE, 95% CI 819-1799 CE).

By combining the analysis of the Cortese *et al*. (2019) haplotype age with our four WGS patients, we were able to estimate the MRCA for the combined cohort. Assuming a correlated genealogy, the mutation arose 1518.5 generations ago (95% CI 65.9-2942.5). Assuming one generation is 20 years, the mutation is 30,380 years old (95% CI 1320-58,860). Assuming 25 years, 37,975 years old (95% CI 1650-73,575). This suggests there was a distant common ancestor for both the Māori/Cook Island and Caucasian patients.

Southern Blot

Southern blot of two affected Cook Island heritage and one Maori patient (CI1 II:1, M2 III:4, M2 V:1 and M5 I:1), showed expanded alleles as described by Cortese *et al.* (2019), with two distinct bands in an individual carrying expansions of different sizes, or one band, or a thick band if the expanded alleles had a similar size. The allele size and number of repeated units are detailed in Table 1. The Southern blot image is shown in Supplementary Fig. 1.

Discussion

Our results demonstrate that the *RFC1* pentanucleotide repeat expansion described by Cortese *et al.* (2019) is responsible for CANVAS in a non-Caucasian population. We also show

a novel configuration of the CANVAS pathogenic pentanucleotide repeat, which appears to be specific to the Māori population. Despite the difference in repeat configuration, there were no apparent phenotypic differences between this cohort and New Zealand European or international cohorts. However, the small sample size (n=13) limits the ability to detect differences.

The rapid childhood onset presentation of M2 V:1 may represent a variant, early onset form of the disease, as seen in other typically late onset repeat expansion disorders such as Huntington's disease (Cronin *et al.*, 2019) or SCA-7 (Benton *et al.*, 1998). M2 V:1 has two of the three clinically defining features of CANVAS: ataxia and bilateral vestibular failure. Co-occurrence of these features is highly distinctive of CANVAS (having excluded Friedreich's ataxia). The normal nerve cross-sectional area seen in this patient has also been previously described in a CANVAS cohort (Pelosi *et al.*, 2017). Although, until we understand fully the pathogenesis of CANVAS and can test for biomarkers of that, the possibility of a second condition cannot be excluded.

We report rapid eye movement sleep behaviour disorder as a feature of CANVAS for the first time, occurring in both our New Zealand European and Māori patients (Table 2). This was so prominent in one European patient that the diagnosis of multiple systems atrophy was initially entertained, but later excluded.

The 0.24 Mb core shared haplotype between our cohort and that of Cortese *et al*. (2019) supports the single origin of the CANVAS disease allele, as suggested by Rafehi *et al.* (2019). There appears to be a relatively recent founder effect in the Māori population. The MCRA of this cohort is estimated to date to ~650-520 years ago (i.e. ~1370-1500 CE). This estimated time period roughly overlaps with the Māori settlement of New Zealand (~1250-1300 CE (Wilmshurst *et al.*, 2008)). If this allele was present early in the Māori settlement of New Zealand, CANVAS may be especially prevalent in individuals of Māori descent. Further

genetic characterisation of the Māori population would be required to define the allele frequency, and epidemiological studies are required to assess the CANVAS disease burden within this population.

Although variation in the *number* of repeated units is common (Paulson and Arbor, 2006), the discovery of this variable repeat *configuration* is highly unusual for a repeat-expansion disease. AAGAG and AGAGG configurations have been described, but it is unknown if they are definitively pathogenic (Akçimen *et al.*, 2019). Cortese *et al*. (2019) suggested the expansion size is correlated with GC content, with the AAGGG configuration allowing a larger expansion to form. The $(AAAGG)_{10\text{-}25}$ expansion preceding the large AAGGG expansion supports this. The 3% of Caucasian patients from Cortese *et al.* (2019) that had an uncharacterised expansion at the *RFC1* disease locus suggest additional pathogenic configurations will be discovered in time. It will be interesting to see if any novel pathogenic repeat expansions at the *RFC1* locus are population specific or share a core haplotype (Cortese *et al.*, 2019; Rafehi *et al.*, 2019).

Acknowledgements

We sincerely thank the patients and their families for their participation in this study. We thank Dr Bharti Morar for her insight on the haplotype analysis. Additional acknowledgement is made to New Zealand neurologists, Drs Neil Anderson, Peter Bergin, Andrew Chancellor, Nick Cutfield, David Hutchinson, Elizabeth Walker, Mark Simpson and Barry Snow all of whom referred patients to this study and without whose clinical acumen it would not have been possible. We also thank Dr David Perry for providing the MRI images.

Funding

SJB is funded by The Fred Liuzzi Foundation (TFLF) (Melbourne, Australia). AC is funded by Medical Research Council (MR/T001712/1), Wellcome Trust (204841/Z/16/Z) and the Inherited Neuropathy Consortium (INC), which is a part of the NIH Rare Diseases Clinical Research Network (RDCRN) (U54NS065712). HH and MMR are grateful to the Medical Research Council (MRC), MRC Centre grant (G0601943), and MMR is also grateful to the National Institutes of Neurological Diseases and Stroke and office of Rare Diseases (U54NS065712) for their support. HH is also supported by Ataxia UK, The MSA Trust, MDUK and The Muscular Dystrophy Association (MDA). MMR is supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. NGL (APP1117510) and GR (APP1122952) are supported by the Australian National Health and Medical Research Council (NHMRC). GR is also supported by a Western Australian Department of Health Future Health's WA Merit Award. This work is funded by TFLF and NHMRC (APP1080587). The funding agencies were not involved in the design, completion, or writing of this study.

Competing interests

The authors have no competing interests to declare.

Supplementary Material

Supplementary Table 1. List of genes on the broad neurogenetic gene panel used for patient M2 V:1.

Supplementary Table 2. Key clinical information for affected patients. d. Deceased; ↓ : Decreased/reduced. ↑ : Increased. AJ: Ankle jerks. Av: Average. BVL: Bilateral vestibular

loss. CMAP: Compound muscle action potential. DBN: Downbeat nystagmus. Hor: Horizontal. LL: Lower limb. N: normal. N/A: not assessed. NCS: Nerve conduction studies. ND: not documented. Onset: age at symptom onset. SNAPs: Sensory nerve action potentials. TG: Tandem gait. UL: Upper limb. Vert: Vertical. vHIT: Video head impulse test. VVOR: Video visually enhanced vestibulo-ocular reflex. *Patient assessed early in disease course due to having affected sibling known to Neurology service.

Supplementary Table 3. Comparison of core haplotype identified from Cortese *et al.* (2019) with patients from this cohort with WGS sequencing data. Marker SNPs are the same as Cortese et al. (2019). Our patients share the 'core' haplotype region identified in Cortese et al. (2019). Beyond that is an additional shared region that is specific to our cohort, suggesting a more recent common ancestor for our patients.

Supplementary Table 4. Haplotyping summary across all patients with next-generation sequencing data. Highlighted regions are homozygous shared haplotypes.

Supplementary Figure 1. Southern blotting of genomic DNA from 5 affected individuals from 2 families and 1 sporadic case. Patients show two discrete or overlapping bands of 8.9 to 13.7 kb. The sample from CI1 II:3 failed due to low DNA quality or quantity. In the control sample (CTRL), one 5-kb band corresponding to the expected size for reference allele $(AAAAG)_{11}$ is observed. Ladders used are DIG-labelled DNA Molecular Weight Marker II (Roche) (LADDER I, left) containing 8 fragments with the following base pair lengths: 125 (not shown), 564 (not shown), 2027, 2322, 4361, 6557, 9416, and 23,130 bp and DIG-labelled DNA Molecular Weight Marker III (Roche) (LADDER II, right) containing 13 fragments with the

following base pair lengths: 125 (not shown), 564 (not shown), 831, 947, 1375, 1584, 1904, 2027, 3530, 4268, 4973, 5148, and 21,226 bp. N.I., sample not included in this study.

Supplementary Figure 2. MRI scans of individual M2 V:1. A. Age 6: Coronal T2 showing preservation of cerebellar mass but with abnormal signal in the superior deep cerebellar white matter which extended into the superior cerebellar peduncles. B. Age 15: Haste Coronal T2 scan showing mild generalised supratentorial and cerebellar volume loss but with particular prominence of the horizontal cerebellar fissure indicating crus 1 atrophy.

Reference list

Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 2002; 30: 97–101.

Akçimen F, Ross JP, Bourassa C V., Liao C, Rochefort D, Gama MTD, et al. Investigation of the RFC1 Repeat Expansion in a Canadian and a Brazilian Ataxia Cohort: Identification of Novel Conformations. Front Genet 2019; 10: 1–5.

Bahlo M, Bromhead CJ. Generating linkage mapping files from Affymetrix SNP chip data. Bioinformatics 2009; 25: 1961–1962.

Beecroft SJ, Yau KS, Allcock RJN, Mina K, Gooding R, Faiz F, et al. Targeted gene panel use in 2249 neuromuscular patients: the Australasian referral center experience. Ann Clin Transl Neurol 2020: 1–10.

Benton CS, de Silva R, Rutledge SL, Bohlega S, Ashizawa T, Zoghbi HY. Molecular and clinical studies in SCA-7 define a broad clinical spectrum and the infantile phenotype. Neurology 1998; 51: 1081–1086.

Cortese A, Simone R, Sullivan R, Vandrovcova J, Tariq H, Yan YW, et al. Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. Nat Genet 2019; 51: 649–658.

Cronin T, Rosser A, Massey T. Clinical Presentation and Features of Juvenile-Onset Huntington's Disease: A Systematic Review. J Huntingtons Dis 2019; 8: 171–179.

Gandolfo LC, Bahlo M, Speed TP. Dating Rare Mutations from Small Samples with Dense Marker Data. Genetics 2014; 197: 1315–1327.

Paulson H, Arbor A. Repeat Expansion Diseases. In: Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine. Springer Berlin Heidelberg; 2006. p. 1627–1627

Pelosi L, Leadbetter R, Mulroy E, Chancellor AM, Mossman S, Roxburgh R. Peripheral nerve ultrasound in cerebellar ataxia neuropathy vestibular areflexia syndrome (CANVAS). Muscle and Nerve 2017; 56: 160–162.

Pelosi L, Mulroy E, Leadbetter R, Kilfoyle D, Chancellor AM, Mossman S, et al. Peripheral nerves are pathologically small in cerebellar ataxia neuropathy vestibular areflexia syndrome: a controlled ultrasound study. Eur J Neurol 2018; 25: 659–665.

Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Auwera GA Van der, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv 2017: 201178.

Rafehi H, Szmulewicz DJ, Bennett MF, Sobreira NL, Pope K, Smith KR, et al. Bioinformatics-Based Identification of Expanded Repeats: A Non-reference Intronic Pentamer Expansion in RFC1 Causes CANVAS. Am J Hum Genet 2019; 105: 151–165.

Szmulewicz DJ, McLean CA, Rodriguez ML, Chancellor AM, Mossman S, Lamont D, et al. Dorsal root ganglionopathy is responsible for the sensory impairment in CANVAS. Neurology 2014; 82: 1410–1415.

Szmulewicz DJ, Waterston JA, Halmagyi GM, Mossman S, Chancellor AM, McLean CA, et

al. Sensory neuropathy as part of the cerebellar ataxia neuropathy vestibular areflexia syndrome. Neurology 2011; 76: 1903–1910.

Taylor JM, Snow BJ, Roxburgh RH, Bergin PS, Chancellor AM, Cleland JC, et al. Autonomic dysfunction is a major feature of cerebellar ataxia, neuropathy, vestibular areflexia 'CANVAS' syndrome. Brain 2014; 137: 2649–2656.

Wilmshurst JM, Anderson AJ, Higham TFG, Worthy TH. Dating the late prehistoric dispersal of Polynesians to New Zealand using the commensal Pacific rat. Proc Natl Acad Sci 2008; 105: 7676–7680.

Figure legends

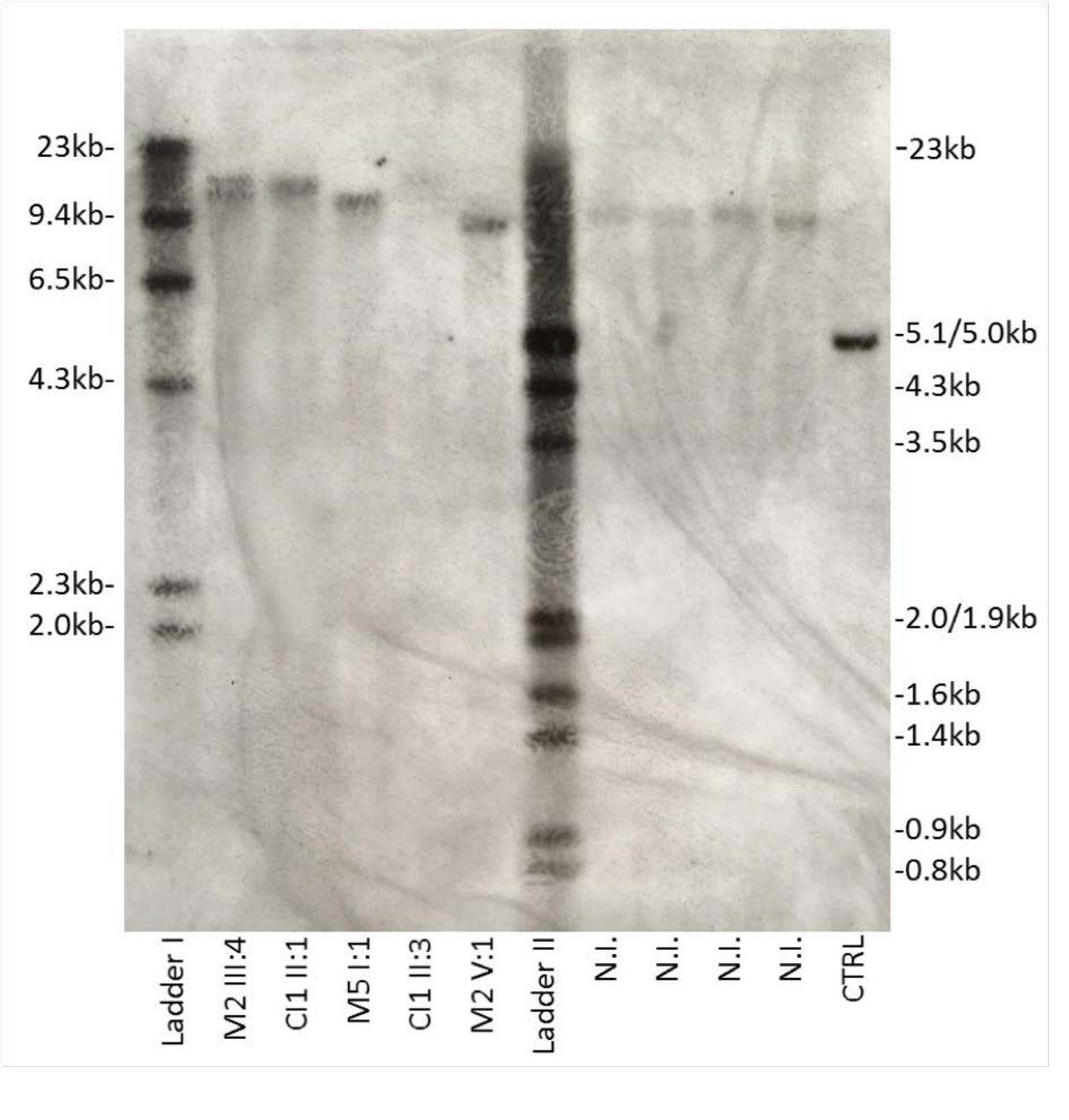
Table 1. Summary of the pedigree and genetic information for the cohort. NGS: Next-generation sequencing. RP-PCR: Repeat primed PCR. WES: Whole exome sequencing. WGS: Whole genome sequencing

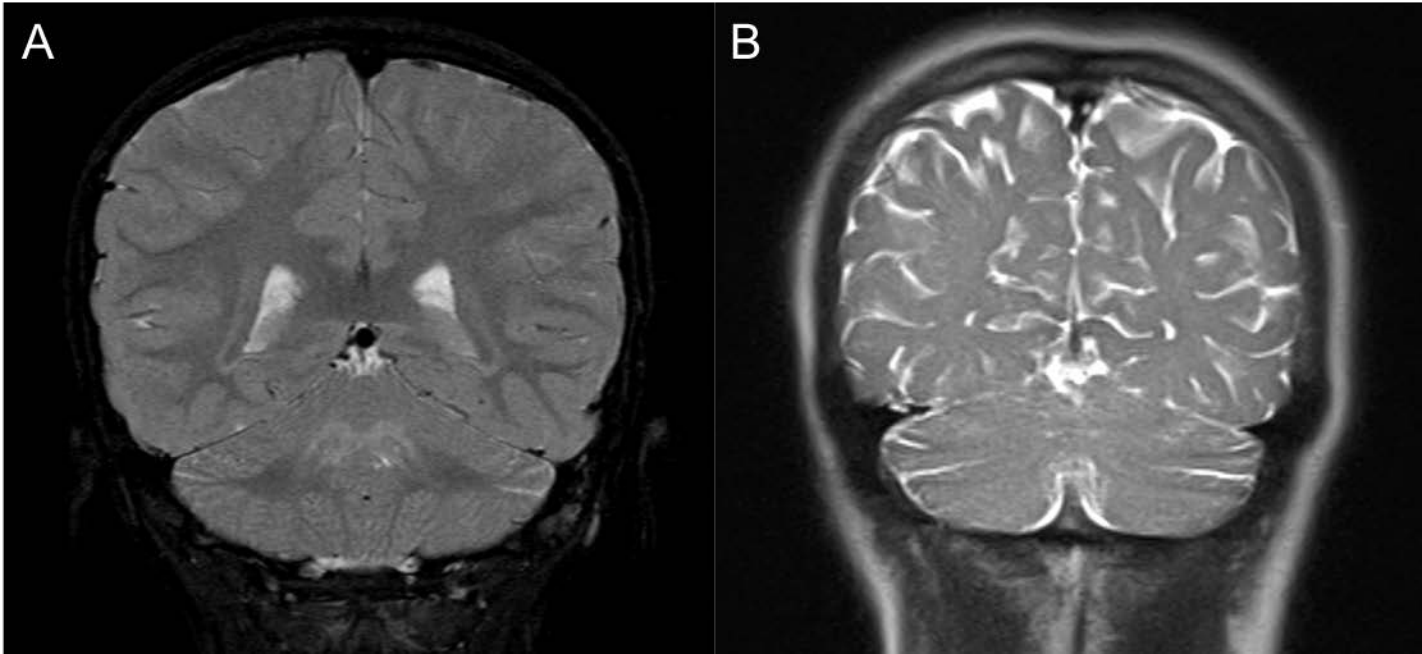
Table 2. Comparison of clinical features of Cook Island and New Zealand Māori patients with the New Zealand European cohort described in Cortese *et al*. 2019 (in press, Brain) and the European cohort described by Cortese *et al.* (2019). For New Zealand cohorts, denominators are the number of patients for whom a symptom or sign is documented. Symptoms refers to patient reported symptoms at any point during the course of illness. Autonomic dysfunction refers to either symptoms of autonomic dysfunction or abnormal autonomic function testing. *Excludes patient M2 V:1. †UK Cohort published by Cortese et *al.* (2019). ‡Oscillopsia or visual blurring or dizziness on head turning. § One additional patient (M2:III:5) had borderline changes.

Figure 1. Pedigrees of the two multiplex families in the cohort. Asterisks indicate individuals that provided DNA for this study. A black circle inside the symbol of an unaffected individual

indicates they are a CANVAS genotype carrier. 1A) Family CI1 comprised Cook Island Māori individuals. 1B) Family M2 comprised New Zealand Māori individuals. Although the first and second generations of family M2 were reported to be unaffected by their family, they were never formally assessed.

Figure 2. Summary of *RFC1* configurations. A) Schematic representation of the repeat alleles at the CANVAS locus, demonstrating the non-pathogenic alleles (top) and the pathogenic alleles (bottom). B) Representative repeat-primed PCR results demonstrating the new CANVAS configuration. The typical *RFC1* $(AAGGG)_{exp}$ repeat pattern is seen in a Caucasian CANVAS patient (top). The novel configuration of $(AAAGG)_{10\text{-}20}(AAGGG)_{exp}$ is shown below. The few (AAAGG) repeats at the distal end of the RFC1 pathogenic expansion were not seen with the repeat-primed PCR. C) Representative soft-clipped reads from WGS BAM file, showing the presence of a small number of AAAGG repeats that are continuous with the distal end of the pathogenic AAGGG repeat.


23kb-
9.4kb-
6.5kb-
4.3kb-
2.3kb-
2.0kb-
-23kb
-5.1/5.0kb
-4.3kb
-3.5kb
-2.0/1.9kb
-1.6kb
-1.4kb
-0.9kb
-0.8kb
Ladder I
M2 III:4
CI1 II:1
M5 I:1
CI1 II:3
M2 V:1
Ladder II
N.I.
N.I.
N.I.
N.I.
CTRL


A
B

Genes on broad neurogenetic panel
AARS
ABCB7
ABCD1
ABHD12
ACTB
ACTG1
ADAR
ADCK3
ADGRG1
AFG3L2
AIFM1
AIMP1
AKT1
AKT3
ALDH18A1
ALDH3A2
ALS2
AMPD2
ANG
ANO10
ANO3
AP4B1
AP4E1
AP4M1
AP4S1
AP5Z1
APTX
ARFGEF2
ARHGEF10
ARL6IP1
ARSA
ARSI
ARX
ASAH1
ASCC1
ASPA
ASPM
ATCAY
ATL1
ATL3
ATM
ATP1A3
ATP2B3

ATP2B4
ATP6AP2
ATP7A
ATR
B3GALNT2
B4GALNT1
BCAP31
BEAN1
BICD2
BSCL2
C12orf65
C19orf12
CACNA1A
CACNA1B
CACNA1G
CACNB4
CCT5
CDK5RAP2
CENPJ
CEP152
CHCHD10
CHMP2B
CLCN2
CLPP
COASY
COL4A1
COL4A2
COX6A1
CPT1C
CSF1R
CTDP1
CUL3
CWF19L1
CYP2U1
CYP7B1
DARS2
DCAF8
DCTN1
DCX
DDHD1
DDHD2
DENR
DHTKD1
DNAJB2
DNAJB5

DNM2
DNMT1
DRD2
DRP2
DST
DYNC1H1
EGR2
EIF2B1
EIF2B2
EIF2B3
EIF2B4
EIF2B5
ELOVL4
ELOVL5
ENTPD1
EOMES
ERLIN1
ERLIN2
EXOSC3
EXOSC8
FA2H
FAM126A
FAM134B
FARS2
FAT3
FAT4
FBLN5
FBXO38
FGD4
FGF14
FIG4
FKRP
FKTN
FLNA
FLRT1
FLVCR1
FTL
FUS
FXN
GAD1
GALC
GAN
GARS
GBA2
GCH1

GDAP1
GFAP
GJB1
GJB3
GJC2
GNAL
GNB4
GOSR2
GRID2
GRM1
HACE1
HARS
HEPACAM
HINT1
HK1
HNRNPA1
HNRNPA2B1
HNRNPUL1
HNRNPUL2
HOXD10
HPCA
HSPB1
HSPB3
HSPB8
HSPD1
IBA57
IFIH1
IGHMBP2
IKBKAP
INF2
ISPD
ITPR1
KARS
KCNA1
KCNC3
KCND3
KCTD13
KIAA0196
KIF1A
KIF1B
KIF1C
KIF21A
KIF2A
KIF5A
KIF5C

KLC2
KLC4
KTN1
L1CAM
LARGE1
LAS1L
LITAF
LMNA
LMNB1
LRSAM1
LYST
MAG
MARS
MARS2
MATR3
MCPH1
MED25
MFN2
MLC1
MORC2
MPZ
MRE11A
MTMR2
MTPAP
MYH14
MYH7
NAGLU
NDE1
NDRG1
NEFL
NGF
NIPA1
NOP56
NOTCH3
NT5C2
NTRK1
OPA1
OPTN
OTUD4
PAFAH1B1
PANK2
PARK2
PAX6
PCNT
PDK3

PDYN
PEX10
PFN1
PGAP1
PIK3R5
PLA2G6
PLEKHG5
PLP1
PMP2
PMP22
PMPCA
PNKP
PNPLA6
POLG
POLR3A
POLR3B
POMGNT1
POMGNT2
POMK
POMT1
POMT2
PPP2R2B
PRDM12
PRKCG
PRKRA
PRNP
PRPH
PRPS1
PRRT2
PRX
PSAP
PTEN
RAB3GAP2
RAB7A
REEP1
REEP2
RELN
RNASEH2B
RNASET2
RNF170
RNF216
RPIA
RTN2
RTTN
RUBCN

SACS
SAMHD1
SEPT9
SBF1
SBF2
SCN11A
SCN1A
SCN9A
SCYL1
SETX
SGCE
SH3TC2
SIGMAR1
SIL1
SLC12A6
SLC16A2
SLC17A5
SLC1A3
SLC20A2
SLC25A46
SLC2A1
SLC33A1
SLC52A1
SLC52A2
SLC52A3
SLC5A7
SLC6A3
SLC9A1
SNX14
SOD1
SOX10
SPAST
SPG11
SPG20
SPG21
SPG7
SPR
SPTBN2
SPTLC1
SPTLC2
SQSTM1
STUB1
SYNE1
SYT14
TAF1

TARDBP
TBCD
TBK1
TBP
TBR1
TDP1
TECPR2
TFG
TH
THAP1
TNFAIP1
TOR1A
TOR1AIP1
TPP1
TREX1
TRIM2
TRIP4
TRPV4
TSEN2
TSEN34
TSEN54
TTBK2
TTPA
TTR
TUBA1A
TUBB
TUBB2B
TUBB3
TUBB4A
TUBG1
UBA1
UBA5
UBQLN2
UNC79
UNC80
USP8
VAMP1
VAPB
VCP
VLDLR
VPS13A
VPS37A
VRK1
WDR45
WDR48

WDR62
WNK1
WWOX
YARS
YWHAE
ZBTB18
ZFR
ZFYVE26
ZFYVE27

Patient details	Pertinent findings from first neurological examination							Progression	Investigations		
Patient Sex Onset	**Gait**	**Dysarthria**	**Limb ataxia**	**Eye movements & head impulse**	**Reflexes**	**Sensation**	**Other**	**Years from onset to disability milestone**	**Brain MRI**	**NCS**	**VHIT**
C1 II:1 F 48	Ataxic	Mild	Bilateral	Saccadic breakdown of pursuit. Ocular dysmetria	N	N	Syncope	Walker 5; Wheelchair 18	Brainstem & cerebellum atrophy 50 yrs	N/A	BVL. Av gain R0.14, L0.15. 63 yrs
C1 II:3 M 45	Ataxic	Severe	Bilateral	Square wave jerks. Broken pursuit	Absent AJ	N	Dizziness, urinary urgency, chronic cough	Walker 10	Cerebellar vermian & hemispheric parenchymal atrophy 51 yrs	↓ SNAPs. Normal motor study	BVL. Av gain R0.23, L0.33. 53 yrs
M2 III:1 M 67	Impaired TG	N	N	N	Absent AJ	↓ foot pinprick	Chronic cough	Independent	N/A	N/A	N/A
M2 III:2 F 49	Ataxic	Severe	UL & LL	Nystagmus on hor. gaze to the right	↓ AJ	ND	Syncope, urinary incontinence	Walker 6; Wheelchair 9; Bedbound 11	N/A	N/A	N/A
M2 III:3 M 57	Impaired TG	Moderate	Mild UL	N	↓ UL & LL	↓ knee vibration sense	Foot pain, chronic cough	Walker 8	Normal 63 yrs	↓ SNAPs. Mildly ↓ LL motor conduction velocities.	BVL. Av gain R0.33, L0.31. Abnormal VVOR 64 yrs
M2 III:4 F 46	Impaired TG	Mild	N	N	Absent AJ	N	Chronic cough	Stick 10; Walker 15	Normal 51 yrs, cerebellar atrophy 62 yrs	Absent SNAPs. ↓ CMAP amplitudes	N/A
M2 III:5 M 59	ND	ND	ND	ND	ND	ND	Hand/foot pain, chronic cough	Independent	N/A	N/A	Catch up saccades. Av gain R 0.67, L0.75. 58 yrs
M2 V:1 F 6	Impaired TG	Moderate	Bilateral	N	Upgoing plantars	N	Numb feet	Wheelchair 2	Dorsal brainstem & cerebellar peduncle T2 hyperintensity 7 yrs	N/A	BVL. Av gain R0.46, L.49. 18 yrs
M3 I:1 M 84	Ataxic	Moderate	Bilateral	DBN on hor. gaze in both directions	ND	↓ knee vibration sense, ↓ toe proprioception	Numb feet, chronic cough	Walker < 1 yr	Cerebellar atrophy 84 yrs	↓ SNAPs. Normal motor study.	BVL. Av gain R0.23, L0.28. 88 yrs
M4 I:1 F 53	Impaired TG	N	Mild UL	DBN in primary position. Abnormal head impulse	N	↓ pinprick & vibration below ankles	Numbness, burning pain	Independent	Mild ventral vermian & Crus I atrophy 54 yrs	↓ SNAPs. Normal motor study.	BVL. Av gain R-0.04, L-0.03. 62 yrs
M5 I:1 M 50	Ataxic	Moderate	N	Rotatory & DBN in all directions. Occasional square wave jerks	↓ UL	↓ vibration LL, ↓ left toe joint position sense	Burning pain in feet, chronic cough	Walker 15	Moderate dorsal & severe ventral vermian atrophy 73 yrs	Absent SNAPs ↓ CMAP amplitudes. Mildly ↓ motor conduction velocities	BVL. Av gain R0.06 L0.04. 77 yrs
M6 I:1 F 54	Ataxic	Mild	UL	DBN on hor. gaze	N	↓ whole limb pinprick, ↓ big toe proprioception	Chronic cough	Walker 4	Cerebellar & Crus I atrophy 56 yrs	↓ SNAPs. Normal motor study	BVL. Av gain R0.4, L0.51. Abnormal VVOR. 51 yrs
M7 I:1 M 56	Impaired TG	Mild	Bilateral	DBN. Positive head impulse	↑ at triceps, ↓ AJ	↓ whole limb vibration	Postural dizziness	Crutches 2; Walker 3	Mild ventral vermian & Crus I atrophy 56 yrs	Absent SNAPs. Normal motor study	BVL. Av gain R0.22, L0.21. 60 yrs

Marker ID	Position	Patient ID				Annotation
		C1 II:1	C1 II:3	M2 III:4	M2 V:1	
rs6814900	chr4:38157510	1/2	1/2	1/1	1/1	
rs2011590	chr4:38944101	1/1	1/1	1/1	1/1	Cohort specific haplotype
rs7694333	chr4:38952030	1/1	1/1	1/1	1/1	Cohort specific haplotype
rs17583377	chr4:38955135	1/1	1/1	1/1	1/1	Cohort specific haplotype
rs6841592	chr4:38965986	1/1	1/1	1/1	1/1	Cohort specific haplotype
rs1864495	chr4:39036311	1/1	1/1	1/1	1/1	Cohort specific haplotype
rs2566120	chr4:39055870	1/1	1/1	1/1	1/2	Cohort specific haplotype
rs2566134	chr4:39074295	2/2	2/2	2/2	2/2	Same extended haplotype as Cortese *et al.*
rs3733275	chr4:39122697	1/1	1/1	1/1	1/1	Same extended haplotype as Cortese *et al.*
rs2044917	chr4:39149491	1/1	1/1	1/1	1/1	Same extended haplotype as Cortese *et al.*
rs2711991	chr4:39151353	1/1	1/1	1/1	1/1	Same extended haplotype as Cortese *et al.*
rs2062229	chr4:39286766	2/2	2/2	2/2	2/2	Same extended haplotype as Cortese *et al.*
rs2066790	chr4:39318706	1/1	1/1	1/1	1/1	Same core haplotype as Cortese *et al.*
rs11096992	chr4:39329102	2/2	2/2	2/2	2/2	Same core haplotype as Cortese *et al.*
rs17584703	chr4:39364856	2/2	2/2	2/2	2/2	Same core haplotype as Cortese *et al.*
rs6844176	chr4:39366590	1/1	1/1	1/1	1/1	Same core haplotype as Cortese *et al.*
rs6823497	chr4:39385821	1/1	1/1	1/1	1/1	Cohort specific haplotype
rs13135439	chr4:39405151	2/2	2/2	2/2	2/2	Cohort specific haplotype
rs11940694	chr4:39414993	2/2	2/2	2/2	2/2	Cohort specific haplotype
rs7674434	chr4:39419409	1/1	1/1	1/1	1/1	Cohort specific haplotype
rs13129975	chr4:39526641	2/2	2/2	1/1	1/1	
rs2381387	chr4:39531928	2/2	2/2	2/2	2/2	
rs10026108	chr4:40379061	2/2	2/2	2/2	2/2	
rs10008483	chr4:40712481	2/2	2/2	2/2	2/2	

Marker ID	Position	C1 II:1	C1 II:3	M2 III:4	M2 V:1	M3 I:1	M4 I:1	M5 I:1	M6 I:1	M7 I:1
rs9995922	chr4:36316348	1/2	1/2	1/1	1/1	1/1	1/1	1/1	1/2	1/1
rs10049643	chr4:36815755	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1
rs2973217	chr4:37430794	1/1	1/1	1/1	1/1	1/2	1/1	1/2	1/2	1/2
rs2276943	chr4:37635128	1/2	1/2	2/2	2/2	2/2	2/2	1/2	2/2	1/1
rs6531590	chr4:37855485	1/1	1/1	1/1	1/1	0/0	0/0	0/0	0/0	0/0
rs2279027	chr4:37902135	1/1	1/1	1/1	1/1	1/1	1/1	1/2	1/2	1/2
rs623945	chr4:38118264	2/2	2/2	2/2	2/2	1/1	2/2	2/2	2/2	2/2
rs2637706	chr4:38510250	2/2	2/2	2/2	2/2	0/0	0/0	2/2	0/0	0/0
rs17616226	chr4:38689805	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2
rs5743810	chr4:38828729	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2
rs2271031	chr4:38935751	2/2	2/2	2/2	2/2	1/2	2/2	2/2	2/2	2/2
rs3733287	**chr4:39215732**	1/1	1/1	1/1	1/1	1/1	0/0	1/1	0/0	1/1
rs4975020	chr4:39527132	1/1	1/1	1/1	1/1	0/0	1/1	1/1	1/1	1/1
rs7688174	chr4:40243362	1/1	1/1	1/1	1/1	0/0	0/0	1/1	0/0	0/0
rs10022491	chr4:40335891	1/2	1/2	2/2	1/2	1/1	2/2	2/2	2/2	2/2
rs278981	chr4:40425993	1/1	1/1	1/1	1/1	1/2	1/1	1/1	1/1	1/1
rs278950	chr4:40502030	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1
rs4861066	chr4:40808646	1/1	1/1	1/1	1/1	1/2	1/1	1/1	1/1	1/1
rs10025369	chr4:41014467	2/2	2/2	1/2	1/1	0/0	0/0	1/2	2/2	0/0
rs17528160	chr4:41263904	1/1	1/1	1/1	1/1	1/1	0/0	1/1	0/0	0/0
rs4266323	chr4:41524486	0/0	1/1	1/1	2/2	1/2	2/2	2/2	2/2	2/2
rs17528897	chr4:41661604	2/2	2/2	2/2	1/1	1/2	2/2	1/2	2/2	1/2
rs12643893	chr4:42067012	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/2
rs898500	chr4:42637169	1/1	1/1	1/2	1/1	1/2	1/1	1/1	1/1	1/2
rs4861050	chr4:43030595	1/1	1/1	1/1	1/1	1/1	1/1	1/1	0/0	1/1
rs10022054	chr4:43899131	1/2	1/2	1/1	1/1	1/1	1/1	1/2	0/0	1/1
rs13117500	chr4:44293822	2/2	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2