Brief Communication

Transcriptome Analysis of a Ring Chromosome 20 Patient Cohort

Kenneth A. Myers*, 1,2,3 Mark F. Bennett,1,4,5 Michael S. Hildebrand,1 Matthew J. Coleman,1 Geyu Zhou,4 Georgie Hollingsworth,1 Anita Cairns,6 Kate Riney,6,7 Samuel F. Berkovic,1 Melanie Bahlo,4,5 Ingrid E. Scheffer1,8,9

Affiliations:
1 Epilepsy Research Centre, Department of Medicine, The University of Melbourne, Austin Health, Heidelberg, Victoria, Australia
2 Research Institute of the McGill University Medical Centre, Montreal, Quebec, Canada
3 Division of Neurology, Department of Pediatrics, Montreal Children’s Hospital, McGill University Health Centre, Montreal, Quebec, Canada

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/EPI.16766

This article is protected by copyright. All rights reserved
Ring chromosomes occur when the ends of normally rod-shaped chromosomes fuse. In ring chromosome 20 (ring 20), intellectual disability and epilepsy are usually present, even if there is no deleted coding material; the mechanism by which individuals with complete ring.
chromosomes develop seizures and other phenotypic abnormalities is not understood. We investigated altered gene transcription as a contributing factor by performing RNA sequencing analysis (RNA-seq) on blood from seven patients with ring 20, and 11 first-degree relatives (all parents). Geographic analysis did not identify altered expression in peri-telomeric or other specific chromosome 20 regions. RNA-seq analysis revealed 97 genes potentially differentially expressed in ring 20 patients. These included one epilepsy gene, NPRL3, but this finding was not confirmed on reverse transcription Droplet Digital PCR analysis. Molecular studies of structural chromosomal anomalies such as ring chromosome are challenging and often difficult to interpret because many patients are mosaic, and there may be genome-wide chromosomal instability affecting gene expression. Our findings nevertheless suggest that peri-telomeric altered transcription is not the likely pathogenic mechanism in ring 20. Underlying genetic mechanisms are likely complex and may involve differential expression of many genes, the majority of which may not be located on chromosome 20.

Key Words: Ring chromosomes; Ring chromosome 20 syndrome; Sequence analysis, RNA; NPRL3; Focal epilepsy; Focal cortical dysplasia.

Introduction

Ring chromosome syndromes are a fascinating genomic phenomena which, despite discovery over 90 years ago, remain poorly understood. Rather than having a full complement of rod-shaped chromosomes, affected individuals have at least one chromosome with circular formation, due to fusion of distal chromosome regions. Ring chromosomes may be supernumerary or individuals may have a normal diploid complement of genetic material; mosaicism also commonly occurs. Though most ring chromosomes arise de novo, parental transmission is possible.

In humans, ring chromosomes are described for all 23 chromosomes. A given ringed chromosome often has a recognizable syndrome; for example, ring chromosome 20 (ring 20) patients have varying degrees of intellectual disability, and a characteristic epilepsy phenotype.
Ring chromosome syndromes were initially thought to occur due to gene deletion on distal chromosome ends; however, this hypothesis does not explain the phenotypic abnormalities seen in individuals with “complete” ring chromosomes, i.e. without detectable deletion of coding genetic material. One previously postulated hypothesis is that abnormal conformation and relative inaccessibility of telomeres results in altered gene expression;\(^6\) however, this has not been tested. Humans with ring chromosomes are rare; the exact prevalence is unknown, but there are only ~60 individuals with ring 20 described in the literature.

We investigated whether phenotypic aspects of ring chromosome syndromes might occur due to altered transcription through analysis of transcript levels in blood with RNA-sequencing (RNA-seq) from a cohort of unrelated ring 20 patients, and comparison to data from healthy controls (unaffected family members). Our bioinformatic approach was designed to answer four main questions:

1. Is gene expression for patients with ring 20 altered at specific regions on chromosome 20 (i.e. near putative breakpoints) when compared to healthy controls?
2. Are there differentially expressed genes in ring 20 patients compared to healthy controls?
3. Is expression of epilepsy genes altered in ring 20 patients when compared to healthy controls?
4. Is there enrichment of differentially expressed genes for specific gene ontologies or biochemical pathways in ring 20 patients compared to healthy controls?

**Methods**

We reviewed the Epilepsy Research Centre’s Epilepsy Genetics Database for patients with ring 20. The degree of mosaicism of ring 20 was estimated for each patient by clinical testing with conventional karyotype analysis studying 15-30 metaphases. We obtained DNA and RNA samples from whole blood from each patient as well as one or two unaffected first-degree relatives using QIAamp DNA Maxi and PAXgene Blood RNA Kits (Qiagen; Hilden, Germany). In all cases, blood samples were collected from patient and relatives at the same time, using the same technique. Informed written consent was obtained from all individuals or their guardians.
The study was approved by the Human Research Ethics Committee of Austin Health, Project No. H2007/02961.

Whole blood samples were prepared for 18 individuals using Illumina TruSeq strand-specific kit, with 75 bp, paired-end reads using the HiSeq4000 or NextSeq500. RNA-seq reads were aligned to the GRCh38.p7 reference genome using STAR (version 2.5.2b). The number of uniquely mapped fragments aligned to each gene was counted using featureCounts using gene annotations from Gencode release 25. Two technical replicates were run for each sample and combined after establishing they were highly correlated (Pearson correlation coefficient applied to log-normalised counts > 0.98 for all replicates).

Statistical analysis to identify differentially expressed genes between ring 20 patients and unaffected family members was performed using the edgeR Bioconductor package. Haemoglobin genes and genes on the sex chromosomes were removed and the remaining 14,316 genes expressed in whole blood RNA-seq data (counts per million > 1 in at least five samples) were included in the differential expression analysis. We fit a generalised linear model using edgeR with covariates to account for sex, separate families, individual age and the estimated degree of mosaicism (Table). Three RUV coefficients were included to remove unwanted variance using the RUV 4-step algorithm with k=3 using a set of 575 housekeeping genes. We accounted for multiple testing by controlling the false discovery rate (FDR) with a threshold of FDR < 0.05 for a gene to be considered significantly differentially expressed.

Gene set enrichment analysis was performed using Fisher’s exact test to determine whether differentially expressed genes were overrepresented in the set of genes from chromosome 20 relative to genes from all other autosomes. Gene sets from Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were also tested for enrichment of differentially expressed genes using the functions implemented in the limma Bioconductor package and KEGG pathway analysis was performed using the Signaling Pathway Impact Analysis method.
For genes associated with epilepsy and found to have significantly altered expression on RNA-seq, we performed follow-up reverse transcription Droplet Digital PCR (ddPCR). NPRL3 expression was normalised using the house-keeping gene, cytochrome C1 (CYC1) at 8q23.4, as an endogenous control. RNA was extracted from whole blood samples and cDNA was synthesised from RNA using the SuperScript III First-Strand Synthesis kit (ThermoFisher, Waltham MA). ddPCR primers were designed across NPRL3 exon 2-3, 7-8 and 10-11 junctions (NM_001243247) (oligonucleotide sequences available on request). Amplification reactions were cycled using a standard protocol on a BioRad C1000 Touch Thermal Cycler (Hercules, CA) annealing temperature of 55°C for 1 minute and absolute quantification was measured on the BioRad QX200 Droplet Digital PCR System using the standard EvaGreen absolute quantification protocol with 10 ng cDNA input. Three technical replicates were run for each sample. CYC1 expression was determined using primer sets designed in-house (sequences also available on request).

All data generated or analysed during this study are included in this published article (and its Supplementary Information files).

**Results**

We obtained DNA and RNA samples from seven individuals (five females) with ring 20, as well as 11 first-degree relatives (all parents of probands). Age of ring 20 patients at time of sample collection ranged from 12 to 44 years.

All patients displayed the epilepsy phenotype and developmental profile consistent with ring 20, albeit with variable severity; 6/7 had ring 20 mosaicism, varying from 7% to 93% (Table). Patients #3-7 were published previously. Prior to her genetic diagnosis, patient #4 had two right parietal lobe resections due to suspected brain malformation. Histopathology revealed mild cortical dyslamination with occasional bare cortical areas and abnormally grouped neurons, but no balloon cells; the pathologic diagnosis was mild cortical dysplasia. She has continued refractory seizures despite these surgical interventions.
**Chromosome 20 Analysis**

After removing genes with low expression in all samples, 399 genes from chromosome 20 remained suitable for analysis. Figure 1A shows the fold change of expression of each gene at its geographic location on the chromosome for each ring 20 patient, compared against their healthy relatives. There was no apparent drop-off in expression for any individual with ring 20, indicating there was no significant abnormality in gene expression at any specific chromosomal loci, including peri-telomeric regions of chromosome 20. Within the resolution of this technique, we found no evidence of a deleted region in any patient, meaning all cases appeared to have functionally “complete” ring 20.

**Differentially Expressed Genes from the RNA-Seq Analysis**

Of the 14,316 genes expressed in whole blood across all autosomes, 97 were differentially expressed in ring 20 patients (Supplementary Data; results for all 14,316 genes are also in Supplementary Data). Of these, one is associated with epilepsy: NPRL3 (OMIM #600928; chromosome 16p13.3), for which expression was decreased 14-fold (FDR = 5.6 x 10^-7), highlighted in the volcano plot (Figure 1B).

However, importantly, decreased expression of NPRL3 was not observed in ddPCR analysis. The difference in NPRL3 expression between ring 20 and controls determined by ddPCR varied not just by family, but also depending on which exons were analysed (Figure 2).

Within the group of differentially expressed genes, there was no overrepresentation of genes from chromosome 20 (p = 0.35), nor was there overrepresentation of epilepsy genes (p = 1).

Covariates were included in the analysis to account for the effect of family, sex, age and degree of mosaicism. The effect of these confounding factors is illustrated using principal component analysis of the 5,000 most highly expressed genes (Supplementary Figure 1). Clustering is initially seen between groups of individuals within families (left panels),

This article is protected by copyright. All rights reserved
however ring 20 patients and unaffected relatives separate after adjusting expression to account for confounding factors (right panels).

**Gene Set Analysis**
The results of gene ontology and KEGG pathway analyses are shown in Supplementary Data. The assessment of specific gene ontology found only that, among differentially expressed genes, there was a non-significant trend towards overrepresentation of genes related to immune system process (FDR = 0.209). The KEGG pathway analysis did not demonstrate any pathways significantly inhibited or activated, after correcting for multiple testing.

**Discussion**
This study of transcriptional patterns in patients with ring 20 was challenging due to both biological and technical reasons, but nevertheless provides some insight into possible pathophysiologic mechanisms that result in syndromic features in these individuals. We did not find any specific altered expression in peri-telomeric regions of chromosome 20, suggesting this may not be the pathogenic mechanism in ring 20. Future studies using independent methods will be required to confirm this.

Our RNA-seq analysis found differential expression of 97 genes; however, we were unable to confirm a key result, downregulation of \textit{NPRL3}, on follow-up ddPCR analysis. The reasons for the inconsistency are not completely clear, but possibilities include differential \textit{NPRL3} transcript expression or the differences in reference genes used. RNA-seq and other similar molecular techniques require normalization of results to a reference or “housekeeping” gene or genes.\textsuperscript{17} Doing this in an appropriate manner is challenging in all experiments, but particularly in a study such as this where widespread altered gene expression with a poorly-understood mechanism is expected. The RNA-seq protocol used a panel of 575 housekeeping genes,\textsuperscript{11} while in ddPCR we used a single well-established gene, \textit{CYC1}.\textsuperscript{15}
There are other reasons to consider our data with caution, as this study has other limitations. Not all genes were expressed at a sufficiently high level in whole blood to allow for analysis; some of our epilepsy gene candidates such as *CHRNA4* and *KCNQ2* which are on the end of 20q near the breakpoint could not be included in the analysis due to low overall expression. Furthermore, we assessed transcriptional patterns in blood, and differential gene expression is likely to be different in brain and other tissues. As well, we collected samples from individuals 12 years and older, and crucial changes in transcription may occur during brain development, a possibility our present study could not accurately assess. We also note that all but one of our patients had mosaic ring chromosome and results may have been more definitive with more non-mosaic (i.e. 100% of cells having ring 20) patients; however, this is essentially unavoidable as in the largest case series of ring 20, 21/28 patients were mosaic.\(^6\) Finally, ring 20 patients were compared to available unaffected parents. As a result, there were a number of confounding factors, which may have influenced the results. Individuals within each family are related and were observed to cluster in family groups (Supplementary Figure 1). Age differed, both within the group of individuals with ring 20 as well as between the ring 20 group and their healthy parents. We accounted for these factors by including covariates in the generalized linear model, however they remain confounding factors that differ between the ring 20 patients and healthy family members, and could have affected our results.

Nevertheless, there were still interesting findings which may lay the groundwork for further study. One of our most important findings is that altered peri-telomeric gene expression is unlikely to be the mechanism of pathology in ring chromosome syndromes. This dogma dates back to the pre-microarray era when many presumed that ring chromosome patients simply had microscopic deletions not appreciated on conventional karyotyping. Whilst RNA-seq provides much finer resolution screen for these events, and immediately assays the transcriptional consequences, genome sequencing, and in particular long-read sequencing, could be performed in the future to help pinpoint precise breakpoints in each patient.

The RNA-seq finding of downregulation of *NPRL3* was initially exciting as mutations in this gene are associated with focal epilepsy, including familial focal cortical dysplasia.\(^{18-20}\) *NPRL3* encodes a subunit of the GATOR complex, which inhibits MTOR, a protein complex that plays a
major role in regulation of cellular proliferation and metabolism. One other MTOR pathway gene, STRADB, also had significantly decreased expression in the RNA-seq analysis (the expression of all MTOR pathway genes is in Supplementary Data). Although the decreased NPRL3 expression was not ultimately validated with ddPCR, the finding of mild cortical dysplasia in patient #4 nevertheless suggests that derangement of the MTOR pathway could partially account for the presence of seizures in patients with ring 20.

In summary, we performed the first RNA-seq analysis of a ring 20 cohort. The analysis of this unique dataset presented challenges not usually encountered in RNA-seq studies. Our findings suggest that the clinical phenotype of ring 20 does not likely occur due to altered expression in peri-telomeric regions. The underlying pathology may relate to complex alteration in expression of many genes, including those not located on chromosome 20; however, delineating these molecular pathways is difficult due to the unique challenge of studying structurally abnormal, often mosaic, chromosomes in human patients.

Acknowledgements
We thank the patients and their families for their participation in this research. This work was supported by the National Health and Medical Research Council [grant number 1091593].

Ethical Publication Statement
We confirm that we have read the Journal’s position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

Disclosure Statement
KAM has received a travel grant from Zynerba, and receives research support from Citizens United for Research in Epilepsy (CURE), Research Institute of the McGill University Health Centre, the Savoy Foundation, and dravetCanada. MSH receives research support from the NHMRC. SFB receives research support from the NHMRC. MB receives research support from the NHMRC, NZ HRC, NIH, March of Dimes, CureKids NZ, Lowy Medical Research Institute, The Walter and Eliza Hall Institute of Medical Research, and the Murdoch Children’s Research Institute. IES serves on the editorial boards of Neurology® and Epileptic Disorders; may accrue

This article is protected by copyright. All rights reserved
future revenue on a pending patent re: Therapeutic compound; has received speaker honoraria from Athena Diagnostics, UCB, GSK, Eisai, and Transgenomics; has received scientific advisory board honoraria from Nutricia and GSK, has received funding for travel from Athena Diagnostics, UCB, and GSK; and receives/has received research support from the NHMRC, ARC, NIH, Health Research Council of New Zealand, March of Dimes, the Weizmann Institute, CURE, US Department of Defense, and the Perpetual Charitable Trustees. The other authors have indicated they have no relevant competing interests to disclose.

Figure Captions
Figure 1 – (A) Geographical Plot of Chromosome 20 Gene Expression in Ring 20 Patients Compared to Controls. The expression of 399 genes on chromosome 20 is plotted for each individual, compared relative to their respective parents. None of the patients demonstrate a point of “drop-off” at any point along the chromosome where expression drops in all genes; this
suggests all seven patients have functionally “complete” ring chromosomes, at least within the resolution of this technique. There is no consistent pattern of regional marked change in gene expression. (B) Volcano plot of differential expression results. The fold change and p-value from differential expression analysis between ring 20 patients and unaffected family members for 14,316 genes. The 97 genes that are differentially expressed with FDR < 0.05 are indicated with blue points and one known epilepsy gene which is differentially expressed, NPRL3, is highlighted in red.

Figure 2 – Results of Droplet Digital PCR NPRL3 Evaluation. The NPRL3 mRNA levels, normalized to CYC1 expression, are shown for primers to exon 2/3 junction (A), exon 7/8 junction (B), and exon 10/11 junction (C). No significant differences were seen, though there was marked variability in relative NPRL3 expression when comparing different exons studied within individual families.

References


This article is protected by copyright. All rights reserved


Table: Clinical and molecular characteristics of the patients.

<table>
<thead>
<tr>
<th>Patient #/Sex/Age</th>
<th>Mosaicism; Breakpoints</th>
<th>Age of Seizure Onset</th>
<th>Seizure Types</th>
<th>Epilepsy Syndrome</th>
<th>Drug-Resistant Epilepsy?</th>
<th>Anticonvulsants at time of collection</th>
<th>Neuropsychiatric Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/M/34 y</td>
<td>100%; complete</td>
<td>9 y</td>
<td>T, A</td>
<td>LGS</td>
<td>Yes</td>
<td>TPM, VPA, LTG</td>
<td>Severe ID. Development was normal to age 9, then regressed with seizure onset. Has episodes of aggression and psychosis.</td>
</tr>
<tr>
<td>2/F/12 y</td>
<td>73%; complete (* 1 cell had 2 r(20), and 1 45(XX)</td>
<td>6 y</td>
<td>FIAS (frontal lobe)</td>
<td>-</td>
<td>Yes</td>
<td>LAC, RUF, PB</td>
<td>Learning difficulties. Intermittent aggression.</td>
</tr>
<tr>
<td>3/F/44 y</td>
<td>7%; p13-q13.33 (estimate from skin fibroblast)</td>
<td>13 y</td>
<td>T, FIAS</td>
<td>-</td>
<td>Yes</td>
<td>CBZ</td>
<td>Mild ID. Development was normal to age 13, then regressed with seizure onset.</td>
</tr>
<tr>
<td>4/F/30 y</td>
<td>60%; p13-q13.33</td>
<td>7 y</td>
<td>FIAS</td>
<td>SHE</td>
<td>Yes</td>
<td>LAC, LEV, LTG</td>
<td>Mild ID. Emotional lability, mood issues.</td>
</tr>
<tr>
<td>5/M/24 y</td>
<td>30%; p13/q13.3</td>
<td>10 y</td>
<td>FIAS</td>
<td>-</td>
<td>Yes</td>
<td>LTG, VPA, ZNS</td>
<td>Mild depression during adolescence. Normal intelligence.</td>
</tr>
<tr>
<td>6/F/35 y</td>
<td>93%; p13-q13.33 (*1 monosomy 20)</td>
<td>8 y</td>
<td>FIAS</td>
<td>-</td>
<td>Yes</td>
<td>CBZ</td>
<td>Moderate ID. Development was normal to age 8, then regressed with seizure onset.</td>
</tr>
<tr>
<td>7/F/28 y</td>
<td>35%; complete</td>
<td>5 y</td>
<td>FIAS</td>
<td>-</td>
<td>Yes</td>
<td>-</td>
<td>Moderate ID. Violent, oppositional, withdrawn.</td>
</tr>
</tbody>
</table>

Abbreviations: A = Absence seizure, CBZ = Carbamazepine, FIAS = Focal impaired awareness seizure, LAC = Lacosamide, LEV = Levetiracetam, LGS = Lennox-Gastaut syndrome, LTG = Lamotrigine, PB = Phenobarbital, RUF = Rufinamide, SHE = Sleep-related hypomotor epilepsy, T = Tonic seizure, TPM = Topiramate, VPA = Valproic acid, ZNS = Zonisamide.

Notes: Mosaicism and breakpoint data is taken from microarrays or karyotypes performed on blood, with the exception of patient #7 whose testing was conducted on skin fibroblasts. Patients #3-7 have been published previously.14
Author/s:
Myers, KA; Bennett, MF; Hildebrand, MS; Coleman, MJ; Zhou, G; Hollingsworth, G; Cairns, A; Riney, K; Berkovic, SF; Bahlo, M; Scheffer, IE

Title:
Transcriptome analysis of a ring chromosome 20 patient cohort

Date:
2021-01

Citation:

Persistent Link:
http://hdl.handle.net/11343/276626