Letter to the Editor:

Copy number variation of the *REXO1L1* gene cluster; euchromatic deletion variant or susceptibility factor?

John CK Barber*[,1] Andrew J Sharp[2], Edward J Hollox[3], Christine Tyson[4]

[1]*Department of Human Genetics and Genomic Medicine, Faculty of Medicine, University of Southampton, Southampton General Hospital, Tremona Road, Southampton, SO16 6YD, UK.*
[2]*Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, 1470 Madison Ave, New York, NY 10029, USA.*
[3]*Department of Genetics, University of Leicester, Leicester, UK.*
[4]*Cytogenetics, Department of Pathology, Royal Columbian Hospital, 330 East Columbia Street, New Westminster, British Columbia, V3L 3W7, Canada.*

*Correspondence: Dr JCK Barber, Department of Human Genetics and Genomic Medicine, Faculty of Medicine, Southampton University School of Medicine, Southampton General Hospital, Tremona Road, Southampton, SO16 6YD, UK. Tel +44 (0)2380 796586; Fax: +44 (0)2380 794760; E-mail: john.barber@soton.ac.uk

Running title: *REXO1L1* Copy number variation

Relatively little is known about the function of the RNA exonuclease 1 homologue (S. cerevisiae)-like 1 gene (*REXO1L1*) in humans and only two titles in PubMed contain the *REXO1L1* acronym. Evidence published in this journal indicated that copy number variation (CNV) of the *REXO1L1* gene and pseudogene cluster was common and likely to be benign[1]. In contrast, it has been proposed that *de novo* heterozygous deletion of the *REXO1L1* cluster in a single patient is responsible for a new microdeletion syndrome including dysmorphic features, cleft palate, incomplete spina bifida, dyspraxia, global developmental delay, growth retardation and gastrointestinal malabsorption[2]. Here we argue that determining the possible clinical significance of large scale copy number variation requires the application of accurate quantitative techniques to both individual families and large cohorts of affected and unaffected individuals.

Microdeletion or euchromatic deletion variant?

The *REXO1L1* gene and pseudogene cluster lies within a 12 kb tandem repeat in band 8q21.2 that is one of the largest Variable Number Tandem Repeat (VNTR) arrays in the human genome[3]. D'Apice *et al*[2] used loss of hybridisation signal from a single Bacterial Artificial Chromosome (BAC) to imply that the whole *REXO1L1* cluster had been deleted but their BAC (RP11-96G1) only partially overlaps this region (Supplementary Figure 1) and residual copies of the 12 kb repeat might not have been detected using FISH. BAC RP11-96G1 is also one of the commonest benign CNVs found using BAC aCGH[4] and subject to loss, gain and inversion in the Database of Genomic Variants (DGV).

The *REXO1L1* gene has no haploinsufficiency likelihood score [Dataset 2][5] and, even if the *REXO1L1* cluster was deleted, the median diploid copy number of the *REXO1L1* gene was 166 and ranged from 97 to 277 copies in 216 controls[1] and 165 HapMap individuals[6] using digital NanoString technology (Supplementary Figure 2). If the patient studied by D'Apice *et al*[2] had half the lowest observed diploid copy number of 97, the remaining copies on the homologous chromosome 8 would be expected to compensate for the loss of the cluster from the chromosome 8 carrying the deletion. Using the less accurate method of real-time quantitative PCR[7] with calibrator plasmids as controls, D'Apice *et al*[2] reported diploid copy numbers that were an order of magnitude lower with 16 to 24 diploid copies in 60 controls and 8 in their deleted patient.

High copy number of the *REXO1L1* cluster can result in a common cytogenetically visible euchromatic variant (EV) that was not thought to be linked to the phenotypes in seven patients with reproductive problems (4), autism (1), a history of intellectual difficulties and cystic fibrosis (1) or congenital anomalies better explained by mosaic trisomy 9 (1)[1]. In all seven reported 8q21.2 EV carriers, FISH signals from the variant chromosomes were approximately twice as large as those on the homologous chromosomes[1]. These simple ~2:1 ratios suggest that the extreme high and low ends of the otherwise continuous *REXO1L1* copy number distribution might be due to large reciprocal duplications and deletions of the *REXO1L1* cluster that are cytogenetically visible as EVs (Supplementary Figure 2). As gains of RP11-96G1 were far more frequent than losses in the healthy populations analysed by Redon *et al*[8], D'Apice *et al*[2] may have found a rare benign euchromatic deletion

variant of 8q21.2 rather than a novel microdeletion. If not due to dosage, any possible effects of a deletion would require alternative mechanisms such as enhancer adoption by genes flanking the *REXO1L1* cluster[7] or the release of other genes from epigenetic control.[6]

Susceptibility factors

There is increasing evidence for a role for recurrent CNVs in infectious disease including lower copy number of the chemokine CC motif ligand 3-like 1 gene (CCL3L1) with susceptibility to chronic HCV infection[7]. As immunological reaction against the human *REXO1L1* gene product (GOR) is a marker of HCV infection, a role for *REXO1L1* copy number in HCV infection was suggested[1]. D'Apice *et al*[2] found increased genetic instability and apoptosis in their patient fibroblasts after treatment with DNA damaging agents and, as *REXO1L1* is a 3'-5' exonuclease member of the DEDD superfamily which might control HBV and RNA virus infections, speculated that low *REXO1L1* copy number might predispose and high numbers protect against viral infection.

The patient of D'Apice *et al*[2] also had malabsorption syndrome with inflammatory infiltrates of the gastrointestinal tract and persistent diarrhoea for which a gluten free diet provided temporary resolution. As other large scale CNVs have been associated with inflammatory diseases[7,9] it is conceivable that low *REXO1L1* copy number might predispose to inflammatory gastrointestinal conditions that include celiac, Crohn's and irritable bowel syndrome. The higher frequency of RP11-96G1 gains than losses[8] and the 0.7 skew against lower copy numbers among

controls[1] suggest that there may be factors which select against low *REXO1L1* cluster copy number.

Conclusion

On the current evidence, it is doubtful that all the phenotypes observed by D'Apice *et al*[2] are due to deletion of the *REXO1L1* gene and pseudogene cluster and more likely that these authors have found a benign EV representing a low extreme of normal copy number variation. Understanding the possible role of this variation in gastrointestinal impairment and viral infection would be aided by more extensive testing of copy number in affected cohorts and controls with accurate quantitative techniques such as digital copy number estimation[1,6] or massively parallel sequencing[10]. Phase and haploid copy number could also be established in multigenerational families including those segregating outliers that could still be detected cytogenetically as EVs in routine clinical laboratories.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

WEB RESOURCES

DGV (Database of Genome Variants): http://dgv.tcag.ca/dgv/app/home

UCSC web browser: http://genome.ucsc.edu/

References:

1. Tyson C, Sharp AJ, Hrynchak M *et al*: Expansion of a 12-kb VNTR containing the REXO1L1 gene cluster underlies the microscopically visible euchromatic variant of 8q21.2*. Eur J Hum Genet* 2014; **22:** 458–463.

2. D'Apice MR, Novelli A, di Masi A *et al*: Deletion of REXO1L1 locus in a patient with malabsorption syndrome, growth retardation, and dysmorphic features: a novel recognizable microdeletion syndrome? *BMC Med Genet* 2015; **16**: 20.

3. Warburton PE, Hasson D, Guillem F *et al*: Analysis of the largest tandemly repeated DNA families in the human genome. *BMC Genomics* 2008; **9**: 533.

4. Whitby H, Tsalenko A, Aston E *et al*: Benign copy number changes in clinical cytogenetic diagnostics by array CGH. *Cytogenet Genome Res* 2008; **123**: 94-101.

5. Huang N, Lee I, Marcotte EM, Hurles ME: Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet* 2010; **6**: e1001154.

6. Brahmachary M, Guilmatre A, Quilez J *et al*: Digital genotyping of macrosatellites and multicopy genes reveals novel biological functions associated with copy number variation of large tandem repeats. *PLoS Genet* 2014; **6**: e1004418.

7. Hollox EJ, Hoh BP: Human gene copy number variation and infectious disease. *Hum Genet* 2014; **133**: 1217-1233.

8. Redon R, Ishikawa S, Fitch KR *et al*: Global variation in copy number in the human genome. Nature 2006; **444**: 444-454.

9. Fernandez-Jimenez N, Castellanos-Rubio A, Plaza-Izurieta L *et al*: Analysis of beta-defensin and Toll-like receptor gene copy number variation in celiac disease. *Hum Immunol* 2010; **71**: 833-836.

10. Nuttle X, Itsara A, Shendure J, Eichler EE: Resolving genomic disorder-associated breakpoints within segmental DNA duplications using massively parallel sequencing. *Nat Protoc* 2014; **9**: 1496-1513.

TITLES AND LEGENDS TO FIGURES

Supplementary Figure 1 A screen shot of *REXO1L1* gene and pseudogene cluster in band 8q21.2 from the UCSC browser (GRCh37/hg19). The large solid red horizontal arrow indicates the variant 284 kb *REXO1L1* region in the Reference genome (86,553,129-86,836,909) and the large dashed horizontal line represents the median 996 kb size of this region estimated from the 166 diploid copies of the 12 kb repeat found by Tyson *et al*[1] in 216 controls. The FISH clones RP11-96G1 and RP11-90G23 are in green with RP11-90G23 manually annotated from BAC end sequence data (base pairs 86,570,255-86,817,480). Note that RP11-96G1 only partially overlaps REPD. The flanking Segmental Duplications (SDs) are represented by black double headed arrows between dotted vertical lines and have been labelled REPP (for REPeat Proximal) and REPD (for REPeat Distal). The sequence gap between the SDs is represented by the black bar labelled GAP. The RefSeq genes are in blue with the tandem *REXO1L1* gene array highlighted by the red arrows underneath. The region contains no non-coding sequences. The DGV tracks show gains (blue), losses (red) and inversions (brown). The UCSC SD track is in gold.

Supplementary Figure 2 A box plot distribution of diploid *REXO1L1* gene copy numbers. On the left, the copy numbers (y-axis) in 216 controls with a median of 166[1]. On the right, circles illustrating the copy numbers of 270 and 265 in a family segregating a cytogenetically visible euchomatic variant of 8q21.2.