The use of pyrosequencing to identify copy number variation of 16p11.2 in euchromatic variant carriers and the normal population

National Genetics Reference Laboratory (Wessex)

Victoria Hall[1], Viv K Maloney[1], Helen White[1], Thomas Liehr[2], Marianne Volleth[3], John CK Barber[1]

1 National Genetics Reference Laboratory (Wessex), Salisbury Hospital NHS Trust, Salisbury, Wiltshire, SP2 8BJ, UK, 2 Institute for Human Genetics and Anthropology, Friedrich-Schiller University, Jena, Germany, 3 Department of Medical Genetics, Eberhard-Karls University, Tuebingen, Germany.

Address for correspondence: Joun Barber, National Genetics Reference Laboratory (Wessex), Salisbury Health Care NHS Trust, Salisbury, Wiltshire, SP2 8BJ, UK. Phone: +44 (0)1722 429080; Fax: +44 (0)1722 338095; e-mail: john.barber@salisbury.nhs.uk

Introduction

Euchromatic variants (EVs) of 8p23.1, 9p12, 9qh/q12, 9q13, 15q11.2 and 16p11.2 can be regarded as the cytogenetically visible tip (Barber, 2005) of an iceberg of copy number variation (CNV) with 1237 CNVs in the Database of Genomic Variants (http://projects.tcag.ca/variation/). They frequently involve segmentally duplicated material that is refractory to sequencing and difficult to quantify accurately.

Pyrosequencing is a real time DNA sequencing technology based on luminometric detection of pyrophosphate released when a nucleotide is incorporated by DNA polymerase. The direct correlation between nucleotide incorporation and light emission makes pyrosequencing technology suitable for a number of quantitative applications, such as determination of allele frequencies, gene copy number and CpG methylation.

Here we have successfully used pyrosequencing of paralogues at 16p11.2 and Xq28 to quantify copy number variation in 16p11.2 in EV carriers and the normal population and show that:
1.pyrosequencing independently corroborates estimates of copy number made using semi-quantitative FISH in variants and in the normal population detected using CGH (by e.g. Sharp et al, 2005).
2.pyrosequencing may be a useful technique for the analysis of individual copy number variants in the search for their possible clinical significance.

FISH

Method

Fluorescence in situ hybridisation (FISH) was carried out with cosmids that hybridise to the proximal 16p11.2 pseudogene cassette and a panel of 1 Mb or 37k cloneset BACs that map to 16p11.2 (www.ensembl.org/Homo_sapiens/cytoview).

Results

Enhanced (enh) signals were seen with the original immunoglobulin heavy chain pseudogene (*IgH)* cosmids used to identify this EV (cos,11, 33 and 98).

In case 1, enhanced signals were also seen with BACs 410P5, 1044J9, 408D2 and 378C4. BACs 1044J9, 408D2 and 378C4 correspond to Locus 251 of the Database of Genomic Variants (http://projects.tcag.ca/variation/). BACs 410P5 and 80F22 are non-contiguous and may represent independent variation at Loci 0341 and 0252.
In case 2, enhanced (enh) signals were seen with BACs 410P5, 408D2, 378C4 and possibly 1044J9. BACs 1044J9, 408D2 and 378C4 correspond to Locus 251 of the Database of Genomic Variants (http://projects.tcag.ca/variation/). BAC 410P5 appeared to be duplicated but is non-contiguous and may represent independent variation at Locus 0341.

The gene and pseudogene content of these regions is illustrated in table 1.

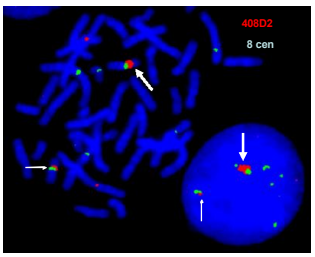

Figure 1: FISH from recent German patient showing contrast between signal strength of BAC from variable region

No	Start (bp)	End (bp)	S	Gene ID and Description (OMIM)	No Ts	K/ N	Gene family (number of members)
Locus 0341 - 410P5; 28072639 - 28390368							
1	28097920	28098045	1	snoRNA gene family member (ENSG00000200682)	1	N	snoACA25 – H/ACA class of snoRNA
2	28102140	28102250	1	ncRNA (ENSG00000200138)	1	N	Y RNA - (component of Ro RNP)
3	28183011	28183181	1	Q81VM4 (ENSG00000198244)	1	K	Unknown
4	28236214	28239743	1	No des (ENSG00000188322)	1	N	Serine/Threonine Kinase 3 (Multiple)
5	28263163	28275423	-1	NP_001018132.1 (ENSG00000196877)	1	K	Unknown
6	28332525	28333958	1	No des (ENSG00000196585)	1	N	HSP90 Chaperone Kinase Targeting Subunit (4)
7	28377001	28386260	-1	NP_001018132.1 (ENSG00000198156)	1	K	Unknown
Locus 0251 - 1044J9-378C4; 31953353 - 32898544							
1	31970812	31971104	1	No des (ENSG00000198565)	1	N	IGHV (20)
2	31977906	32028444	1	XP_372632.2 - similar to IGVH V-III region VH26 precursor (ENSG00000182414)	5	K	IGVH (20)
3	32172144	32188269	1	NP_056184.1 (ENSG00000180398)	2	K	TP53TG3 (4)
4	32297917	32298822	-1	No des (ENSG00000197605)	1	N	Actin 3 related (7)
5	32357891	32358321	-1	No des - pseudogene (ENSG00000197184)	1	N	
6	32579144	32580554	-1	NP_057296.1 (ENSG00000183402)	2	K	TP53TG3 (4)
7	32746520	32746971	1	Similar to IGVH V-III region VH26 precursor (ENSG00000198662)	1	K	IGHV (20)
8	32796298	32803064	-1	SLC6A10 (NP_942154) (Testis) Sodium- and chloride-dependent creatine transporter 2 (CT2) (ENSG00000189172) expressed in testis only (OMIM 601294)	2	K	Sodium dependent transporter CT2 (1) - CT1 maps to paralogous region of Xq28
9	32835792	32866511	1	No des (ENSG00000181601)	1	N	GTPase activating GAP CEGAP (5)
Locus 0252 - 80F22; 34466093 - 34637143							
1	34538908	34539347	-1	No des - pseudogene (ENSG00000188928)	1	N	None
2	34569286	34572357	-1	Low complexity peptide LOC_146481 (OTTHUMG00000073282)	2	V	
3	34635158	34635593	-1	No des – Pseudogene(ENSG00000183290)	1	N	None

Table 1: genes within copy number variant regions of 16p11.2

Pyrosequencing

Method

Pyrosequencing was carried out to quantify the copy number variation in EV carriers and the normal population. Paralogous sequences located at 16p11.2 and Xq28 were coamplified and paralogous sequence mismatches (PSM) between them were quantified using pyrosequencing. The resultant PSM allele frequencies reflect the relative frequency of the paralogous sequence on each chromosome.

Results

Figure 2 shows pyrosequencing traces for a normal male, normal female, and a male and a female patient with grater than 10 copies of 16p. The relative frequency of the paralogous sequences is calculated using the Pyrosequencing AQ software.

To date we have performed pyrosequencing on 12 cases previously analysed by FISH and all results have been comparable, with results ranging from 3 copies to >15 copies

Results from a group of normal controls (20 males and 26 females) indicate that copy number in the normal population ranges from 3 to 8, with a median value of 4, and copy number in EV carriers is greater than 10.

Our results demonstrate that pyrosequencing could be used to assess whether individual copy number variants have any effect on variable phenotypic traits.

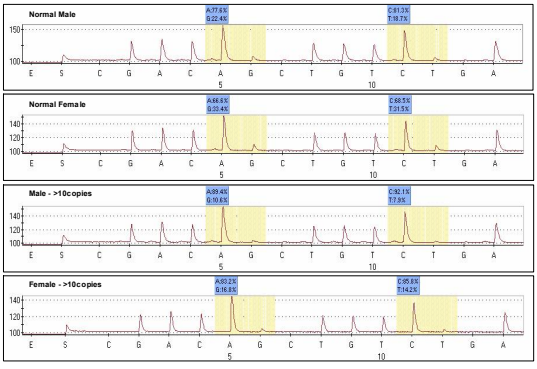

Figure 2: Pyrograms from a Normal male (a), normal female (b) a male with greater than 10 copies of 16p and a female with greater than 10 copies (c).

Discussion

In our study pyrosequencing independantly corroborates with estimates of copy number made using semi-quantitative FISH in varients and in the normal population. In the Database of Genomic Variants there are at least 1,237 copy number varients and many are phenotypically silent. Pyrosequencing could be a useful high throughput technique to look for and verify these copy number varients and could be used to asses whether individual copy number varients have any effect on variable phenotypic traits.
The 16p11.2 EVs identified to date clearly do not have the phenotypic consequences associated with unbalanced chromosome abnormalities (UBCAs). However, their gene content and copy number variation in normal individuals does not exclude a possible role in traits which show continuous variation. It is also interesting that some of the human EVs involve genes that have testis specific expression e.g. TP53TG3 in 16p11.2 EVs and SPAG11 in 8p23.1 EVs. Additional copies of a variable domain might be under strong selection if they conferred a significant effect on fertility.

References:

Barber JCK: Directly transmitted unbalanced chromosome abnormalities and euchromatic variants. *J Med Genet* 2005;42:609-629

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. Large-scale copy number polymorphism in the human genome. Science. 2004;305:525-8.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. Nat Genet 2004 Aug 1 [Epub ahead of print]

Eichler EE, Lu F, Shen Y, Antonacci R, Jurecic V, Doggett NA, Moyzis RK, Baldini A, Gibbs RA, Nelson DL. Duplication of a gene-rich cluster between 16p11.1 and Xq28: a novel pericentromeric-directed mechanism for paralogous genome evolution. Hum Mol Genet 1996;5:899-913

Eichler EE, Budarf ML, Rocchi M, Deaven LL, Doggett NA, Baldini A, Nelson DL, Mohrenweiser HW. Interchromosomal duplications of the adrenoleukodystrophy locus: a phenomenon of pericentromeric plasticity. Hum Mol Genet 1997;6:991-1002