

RESEARCH ARTICLE

Open Access

# miRNA and miRNA target genes in copy number variations occurring in individuals with intellectual disability

Ying Qiao<sup>1,2†</sup>, Chansonette Badduke<sup>1†</sup>, Eloi Mercier<sup>3</sup>, Suzanne ME Lewis<sup>2</sup>, Paul Pavlidis<sup>3</sup> and Evica Rajcan-Separovic<sup>1\*</sup>

## Abstract

**Background:** MicroRNAs (miRNAs) are a family of short, non-coding RNAs modulating expression of human protein coding genes (miRNA target genes). Their dysfunction is associated with many human diseases, including neurodevelopmental disorders. It has been recently shown that genomic copy number variations (CNVs) can cause aberrant expression of integral miRNAs and their target genes, and contribute to intellectual disability (ID).

**Results:** To better understand the CNV-miRNA relationship in ID, we investigated the prevalence and function of miRNAs and miRNA target genes in five groups of CNVs. Three groups of CNVs were from 213 probands with ID (24 *de novo* CNVs, 46 familial and 216 common CNVs), one group of CNVs was from a cohort of 32 cognitively normal subjects (67 CNVs) and one group of CNVs represented 40 ID related syndromic regions listed in DECIPHER (30 CNVs) which served as positive controls for CNVs causing or predisposing to ID. Our results show that 1). The number of miRNAs is significantly higher in *de novo* or DECIPHER CNVs than in familial or common CNV subgroups ( $P < 0.01$ ). 2). miRNAs with brain related functions are more prevalent in *de novo* CNV groups compared to common CNV groups. 3). More miRNA target genes are found in *de novo*, familial and DECIPHER CNVs than in the common CNV subgroup ( $P < 0.05$ ). 4). The MAPK signaling cascade is found to be enriched among the miRNA target genes from *de novo* and DECIPHER CNV subgroups.

**Conclusions:** Our findings reveal an increase in miRNA and miRNA target gene content in *de novo* versus common CNVs in subjects with ID. Their expression profile and participation in pathways support a possible role of miRNA copy number change in cognition and/or CNV-mediated developmental delay. Systematic analysis of expression/function of miRNAs in addition to coding genes integral to CNVs could uncover new causes of ID.

**Keywords:** Micro RNA (miRNA), Copy number variants (CNVs), Copy number variant regions (CNVRs), Intellectual disabilities (ID), Functional pathways

## Background

MicroRNAs (miRNAs) are an abundant class of short, non-coding, endogenous RNAs that regulate gene expression at the post-transcriptional level [1,2]. The mature miRNA is a ~20-23 nucleotides long single stranded sequence, which derives from primary transcript miRNA (pri-miRNAs) [3]. One miRNA can bind to hundreds of

target genes at the 3'-UTR of mRNAs (miRNA target genes), and a single miRNA target gene can be targeted by multiple miRNAs [4,5]. Based on bioinformatic predictions, up to 90% of human genes are believed to be regulated by miRNA [5]. MiRNAs have primarily been demonstrated to mediate translational repression or target gene degradation and silencing. Recently, it has been shown that they can also upregulate gene expression by targeting gene regulatory sequences [6].

Genetic polymorphisms in miRNA or their targets can add to the complexity of miRNA regulation and function. Single nucleotide polymorphisms (SNPs) in miRNA

\* Correspondence: [eseeparovic@cw.bc.ca](mailto:eseeparovic@cw.bc.ca)

†Equal contributors

<sup>1</sup>Department of Pathology and Lab Medicine, BC Child and Family Research Institute, University of British Columbia (UBC), Vancouver, BC V5Z 4H4, Canada

Full list of author information is available at the end of the article

binding sequences have indeed been shown to affect miRNA-mediated gene regulation and alter the expression of target genes [7]. A low SNP density at miRNA genes exists compared to the reference human genome, suggesting a negative selection for miRNA with SNPs [8-10]. CNVs, as a major class of genomic variations, also have an effect on miRNA, as demonstrated by underrepresentation of miRNA in highly polymorphic CNVs compared to the reference genome [10]. In contrast, the number of miRNA target genes in polymorphic CNVs is higher than in non-CNV regions, suggesting that genes integral to polymorphic CNVs are more likely to be regulated by miRNAs, in order to counteract their expression changes due to copy number variability of the region in which they reside [9]. Multiple cancer studies show that miRNAs integral to CNVs demonstrate gain or loss at the genomic level, and are associated with expression changes for ~10% -20% miRNAs [11,12].

In addition to cancer, miRNAs are involved in other human diseases, for example, cardiovascular diseases and neurological/neurodevelopmental disorders [13,14]. Evidence for the role of miRNA in these diseases is based on identifying mutations or differential expression of specific miRNAs and/or their global expression [15-17], and the effect of genomic copy number change on miRNA function is largely unknown. In rare instances association of CNVs with miRNA expression was studied in subject with cognitive delay. For example in Down syndrome an extra copy of chromosome 21 was associated with upregulation of several miRNAs from this chromosome, while their targets are downregulated [18]. In addition, a deletion of 1p21.3 containing a miRNA MIR137 has recently been reported by Willemsen et al. [19], and resulted in downregulation of the miRNA, and upregulation of three targets in subjects with ID and congenital abnormalities. Considering the role of MIR137 and its targets in brain function (synapse maturation, morphogenesis of young neurons, and axon growth), their dysfunction is considered causative of the patients' phenotype [19]. Finally, microdeletion of a miRNA cluster MIR17HG on chromosome 13 is associated with Feingold syndrome which includes microcephaly, skeletal abnormalities and variable levels of ID [20]. The role of this miRNA in the human phenotype is further supported by a mouse knock-out model [20].

Although the above examples demonstrate the relevance of copy number change on miRNA function and its role in human disease, these studies are focusing on one type of CNV exclusively; for example, on common CNVs from normal populations [8-10], lump of CNVs from single disease cohorts such as cancer [11,12] or autism [21], or a few individual pathogenic CNVs associated with neurodevelopmental disorders [19,20]. Considering that ~70% of miRNAs are expressed in brain [22], and function in neurodevelopment, neurotransmission, synaptic plasticity,

neurite outgrowth and dysregulation [18,23], we aimed to characterize and understand the total presence and functions of miRNAs in CNVs detected in idiopathic ID in comparison to neurotypical controls. Five different CNV subgroups were studied for their miRNA and miRNA target gene contents: (i) *de novo*, (ii) familial, (iii) common CNVs detected in subjects with idiopathic ID, (iv) common CNVs in cognitively normal subjects, and (v) CNVs associated with syndromic ID selected from DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources) database (<http://decipher.sanger.ac.uk/>). CNVs from DECIPHER were considered as positive controls. Our study is unique in serving as the first report comparing the miRNA content and function in different CNV subgroups from ID subjects and cognitively typical controls.

## Results

### CNV detection and sub-classification

Initially, we identified 24 *de novo*, 46 familial and 216 common CNVs from 213 cases with idiopathic ID, 67 common CNVs from 32 cognitively normal subjects, and 30 CNVs collected from 40 ID-related syndromic regions in DECIPHER database representing CNVs known to cause or predispose to developmental delay. All the CNVs were identified experimentally except the CNVs from DECIPHER which were retrieved from the DECIPHER database. Overlapping CNVs from the same subgroups were merged into CNVRs (CNV Regions). The final set included 22 *de novo*, 46 familial, 30 DECIPHER CNVRs, and respectively 210 and 61 common CNVRs from cases versus controls (Table 1). The genomic coverage (Build 36 hg18) for all CNVRs and the miRNAs present in each CNVR are provided in Additional file 1: Table S1-S5.

### Characteristics of miRNA in different CNV subgroups

#### Number of miRNAs in different CNV groups

The number and loci of miRNAs in different CNV subgroups were obtained using Galaxy (<https://main.g2.bx.psu.edu/>) and miRBase (<http://www.mirbase.org/>). To avoid bias caused by varying CNV size, the number of miRNA in each CNV region was weighted by the size of the CNV, and expressed as miRNA/Mb. The weighted median number of miRNAs/Mb in each CNV subgroup is shown in Table 1 and was compared among different pairs of CNV subgroups using the Wilcoxon rank-sum test. We found the median number of miRNAs/Mb is significantly higher in *de novo* and DECIPHER CNVs than familial or common CNVs from cases, and common CNVs from controls ( $P < 0.01$ ) (Table 1 and Figure 1). However, the average miRNA/Mb between the common and pathogenic subgroups is comparable which is likely due to the presence of few very small common CNVs with

**Table 1 Comparison of the number of miRNAs in different CNV subgroups**

CNV category	No. of CNVs	No. of CNVRs	Size of total CNVRs (Mb)	Average size of CNVRs (Mb)	miRNA			
					Total No. (#)	Average miRNA/Mb	#/CNVR	Weighted median No.
<i>De novo</i>	24	22	70.5	3.21	84	1.20	3.8	0.6 <sup>a</sup>
Familial	46	46	29.5	3.34	14	0.47	0.3	0 <sup>b</sup>
Common from cases	216	210	105.2	0.64	67	0.64	0.3	0 <sup>b</sup>
Common from control cohort	67	61	34.1	0.56	30	0.88	0.5	0 <sup>b</sup>
Pathogenic from DECIPHER	35	30	100.1	0.50	90	0.90	3.0	0.8 <sup>a</sup>

<sup>a,b</sup>Significant difference was found between data a and b ( $P < 0.05$ ) (a Wilcoxon signed-rank test).

high miRNA density (for example, 8 miRNAs within one common CNV at 16p13 (16,322,499-16,682,499).

#### Genomic coverage of miRNAs in different CNV groups

The genomic coverage of miRNA and protein coding genes (expressed as their number per CNV group per their total number in the whole genome) tended to be higher for *de novo* and DECIPHER CNVs, in comparison to the coverage of miRNAs and protein coding genes of randomly sampled sections from the reference genome of the same size as each of the CNV subgroups. This increase in comparison to the reference genome was the highest for miRNAs in *de novo* CNVs, but was not significant ( $p > 0.05$ , Wilcoxon signed-rank test) (Figure 2). The *de novo* and DECIPHER CNV coverage of protein coding genes was also slightly higher than expected by chance but not significant. In contrast, the protein coding genes covered a significantly smaller fraction of the common CNVs than expected by chance, i.e. when reference genome of the same size as CNVs was used ( $P < 0.01$ ) (Figure 2B), in keeping with their suspected benign nature. For the familial CNVs, the observed-to-expected miRNA and protein coding gene coverage was comparable.

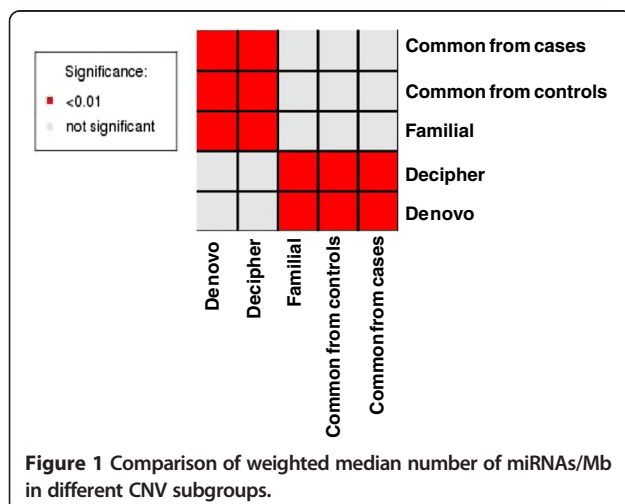
#### Functional characteristics of miRNA from different CNV groups

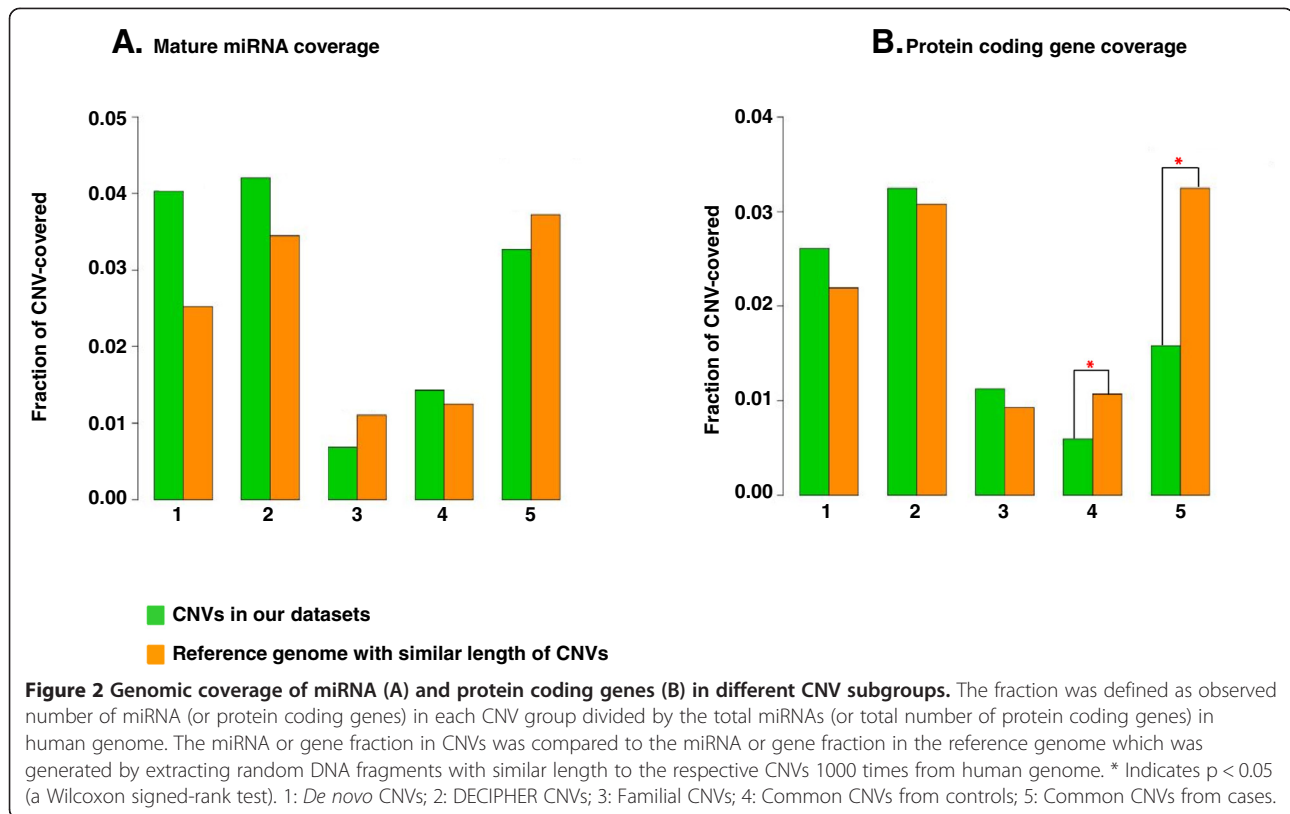
##### Expression of miRNAs from CNVs

The expression profiles of miRNAs in different CNV subgroups were obtained using two web-tools, the miRNA database (<http://mimirna.centenary.org.au/mep/formulaire.html>) [24] and microRNA.org-Targets (<http://www.microrna.org/microrna/home.do>) [25]. Both web-tools identified a small number of miRNAs (20 ~ 30% of miRNAs in *de novo* or familial CNV groups, and ~10% of miRNAs in common CNV groups) with published expression profiles. Although the low number of miRNA with expression/function data does not allow definitive conclusions, we note that the miRNAs expressed in brain related tissues and/or having experimental evidence of involvement in nervous system dysfunction (based on manual search of PubMed) seem to be more prevalent in the *de novo*, familial and DECIPHER CNV group (~50% of all *de novo*, familial and DECIPHER CNV miRNAs with expression data) than in the common CNVs tested (0-25% of all common CNV miRNAs with expression data) (Table 2). A list of miRNAs with available expression data and brain or nervous system related function is shown in Table 3. In comparison to the common CNVs from patients, the two fold increase in the number of brain function related miRNA in familial CNV group suggests that some genes in familial CNVs could contribute to the disease process. There were no brain function-related miRNAs in common CNVs from cognitively typical controls (Table 2).

##### Functional and pathways enrichment analysis of miRNA targets

Using the web-tool WebGestalt, we searched for pathway enrichment of miRNA target genes from different CNV subgroups compared to the reference genome. We found that approximately 8-11% of genes from *de novo* CNVs, DECIPHER CNVs or familial CNVs were targeted by miRNAs in comparison to 0-1% of genes from common CNV subgroups ( $P < 0.01$ ; two-sided Fisher's exact test) (Table 4). KEGG pathway enrichment analysis was performed using WebGestalt for identification of pathways





enriched for the miRNA target genes in different CNV groups. The top 10 pathways are listed in Table 4 and contained nervous system-related pathways (Axon guidance and Neurotrophin signaling, respectively) for only *de novo* and DECIPHER CNVs (Table 4). We found that targets from both *de novo* and DECIPHER CNVs participate in MAPK signaling pathways more than

expected by chance. Ubiquitin-mediated proteolysis was the only pathway found to be enriched for miRNA target genes in the familial CNV subgroup. There was no apparent pathway enrichment for the miRNA target genes from our common CNV subgroup. The miRNAs targeting the genes from the top 10 pathways for each of our CNV subgroups are also listed in Table 4.

**Table 2 Expression/Function of miRNAs in different CNV groups**

CNV category	No. of miRNA	No. of miRNA with available expression data*	Brain expression/function related miRNA**
<i>De novo</i>	84	18	10 (55%)
Familial	14	4	2 (50%)
Common from case cohort	67	4	1 (25%)
Common from control cohort	30	3	0
Pathogenic from DECIPHER	90	24	11 (46%)

Notes: \*indicates data collected from websites: <http://www.microrna.org/microrna/> and [mimiriNa \(http://mimiriNa.centenary.org.au/mep/mir.htm\)](http://mimiriNa.centenary.org.au/mep/mir.htm). \*\*indicate manual PubMed search for each miRNA.

## Discussion

MiRNA-mediated gene regulation is complex and genomic variations, such as CNVs and SNPs, which can modulate miRNA expression, add to this complexity. The miRNA-CNV relationship has been rarely studied and predominantly involved polymorphic CNVs found in control cohorts [9,10].

Our study is novel in its comparison of miRNA content in different classes of CNVs detected in a cohort of subjects with ID relative to cognitively typical subjects. We found a significant increase in the number of miRNAs in *de novo* and DECIPHER CNVs versus common CNVs (weighted median 0.6 and 0.8 versus 0.0,  $P < 0.05$ ). In addition, the miRNAs in *de novo* CNVs were more likely to have expression in brain-related tissues or cell lines compared to the miRNAs from common CNV groups. Our collective findings suggest that miRNAs from *de novo* and putatively pathogenic CNVs could contribute

**Table 3 MiRNAs with expression and/or function related to brain or nervous system**

#Chr	Start (bp)	End (bp)	miRNA ID	CNV feature	Expression*	Functional relevance**	Reference
1	1103258	1103279	hsa-miR-200a/b	<i>de novo</i>	hsa-pancreatic islets, hsa-breast adenocarcinoma, HT29, breast malignant tumor	<b>olfactory neurogenesis; neuronal differentiation of neural stem/progenitor cells</b>	PMID:18184563; PMID: 22993445
1	1104435	1104456	hsa-miR-429	<i>de novo</i>	cancer-related	non-brain cancer related; <b>neuroprotective</b> effect in in vitro ischemia	PMID: 21684154; PMID: 20576953
5	87962684	87962705	hsa-miR-9	<i>de novo</i>	<b>brain, astrocytoma, neuroblastoma</b>	Upregulation in HD (Huntington's disease); involved in <b>spinal motor neuron disease</b> ; increased in <b>fmr1/foxr2 knock-out mice</b>	PMID: 19118166; PMID: 20616011 PMID:21957233
5	179442361	179442382	hsa-miR-340	<i>de novo</i>	<b>neuroblastoma</b>	cancer-related including <b>neuroblastoma</b>	PMID: 22797059
7	99691233	99691253	hsa-miR-25	<i>de novo</i>	<b>neuroblastoma</b>	<b>neural</b> stem cells differentiation	PMID: 21386132
7	99691438	99691460	hsa-miR-93	<i>de novo</i>	cervix-Hela	<b>neural</b> stem cells differentiation	PMID: 21386132
7	99691625	99691646	hsa-miR-106b	<i>de novo</i>	<b>Neurobl</b> -SHSY5Y_IFN, AML-THP1, kidney-embryo-HEK2	<b>neuronal differentiation</b> ; pathogenesis of Alzheimer's diseases	PMID: 21386132; PMID: 20709030
19	4770712	4770734	hsa-miR-7	<i>de novo</i>	<b>neuroblastoma</b>	Repression of alpha-synuclein accumulation in <b>Parkinson's disease</b> ; brain cancers	PMID: 20106983; PMID: 21912681
20	61151558	61151579	hsa-miR-1	<i>de novo</i>	heart, thyroid, Ewing-sarcoma	muscle development; modulating <b>neurite outgrowth</b>	PMID: 22365735; PMID: 21170745
20	61809866	61809887	hsa-miR-124	<i>de novo</i>	<b>hsa-hippocamp-adult, hsa-celebellum-adult</b>	<b>neuronal differentiation and incorporation of neural-specific exons</b> ; increased in <b>fmr1/foxr2 knock-out mice</b>	PMID: 17679093 PMID:21957233
1	94312436	94312455	hsa-miR-760	familial	hsa- <b>Neurobl</b> -SHSY5Y, hsa-Hodgkin-KMH2, hsa- <b>midbrain</b> -adult	non-brain cancer related	PMID: 22970209
15	45725296	45725317	hsa-miR-147b	familial	hsa-fibrobl-CMV, hsa-medullobl-DADY, hsa-DLBLEL-DLBLEL, <b>dendritic cell</b>	rectal cancer-specific	PMID: 22850566
2	32757280	32757298	hsa-miR-558	common-case	ovary and ovary-related cancer	cancer-related including <b>neuroblastoma</b>	PMID: 21498633

Notes: \*indicates data collected from websites: <http://www.microrna.org/microrna/> and mimiRNA (<http://mimrna.centenary.org.au/mep/mir.htm>).

\*\*indicate manual PubMed search for each miRNA.

to ID etiopathogenesis, in addition to coding genes integral to CNVs.

Similar to the increase in the number of miRNAs in *de novo* and DECIPHER CNVs, the number of miRNA target genes integral to these CNV types was also higher in comparison to familial or common CNVs. Lower numbers of miRNAs and miRNA target genes in common CNVs compared to the *de novo* or pathogenic CNVs suggests that they participate in processes that can tolerate functional variation. Previous studies have shown that targets for miRNAs from polymorphic CNVs tend to participate in environment-orientated processes including stimulus responses and immune responses, while the targets from non-CNV regions of the genome are enriched for fundamental biological processes such as maintenance of chromatin, chromosome segregation and nucleic acid processes [26]. In our

study, genes from common CNVs that are targets for miRNAs do not show enrichment in any pathway while those from familial CNVs show enrichment in ubiquitin mediated proteolysis. In contrast, brain function related pathways such as axon guidance, Huntington's disease, neutrophin signaling pathway, are found to be among the top 10 pathways enriched for targets from *de novo* and pathogenic CNVs. Targets from these two CNV groups also showed enrichment in the MAPK (mitogen-activated protein kinase) signaling pathway. The MAPK signaling cascade is involved in a wide variety of cellular processes and has been recently reported to be involved in ID pathogenesis [27-29]. Seven genes from *de novo* and DECIPHER CNVs which are targets for miRNAs were found to be involved in the MAPK pathway (*MAPK9*, *MAPT*, *MEF2C*, *MKNK2*, *PAK2*, *RPS6KA1*, and *TAOK2*). Three of them are known to be related to

**Table 4 Summary of miRNA target genes within the CNV subgroups and the miRNAs targeting the CNV genes**

CNV category	No. of coding genes in CNVs	Genes in CNVs targeted by miRNA	Pathways enriched for miRNA target genes	miRNA targeting the genes in CNVs in enrichment analysis
De novo from ID	737	84 (11.5%) <sup>a</sup>	MAPK signaling pathway Lysosome Insulin signaling pathway Axon guidance Huntington's disease Hedgehog signaling pathway Endocytosis N-Glycan biosynthesis p53 signaling pathway Progesterone-mediated oocyte maturation	MIR-128A,MIR-128B,MIR-194,MIR-27A,MIR-27B,MIR-296,MIR-302A,MIR-302B,MIR-302C,MIR-302D,MIR-326,MIR-329,MIR-34A,MIR-34C,MIR-372,MIR-373,MIR-449,MIR-503,MIR-520A,MIR-520B,MIR-520C,MIR-520D,MIR-520E,MIR-526B,MIR-9,MIR-93
Pathogenic from Decipher	838	87 (10.4%) <sup>a</sup>	Neurotrophin signaling pathway Renal cell carcinoma MAPK signaling pathway Pathways in cancer Chronic myeloid leukemia Chemokine signaling pathway Long-term potentiation ErbB signaling pathway TGF-beta signaling pathway Adherens junction	MIR-196A,MIR-196B,MIR-24,MIR-320,MIR-506,MIR-493,MIR-25,MIR-32,MIR-92,MIR-363,MIR-367,MIR-512-5P,MIR-302C,MIR-30A-5P,MIR-30C,MIR-30D,MIR-30B,MIR-30E-5P,MIR-485-3P
Familial from ID	188	15 (8.1%) <sup>a</sup>	Ubiquitin mediated proteolysis	MIR-193A,MIR-193B,MIR-495,MIR-302C,MIR-198
Common from ID case cohort	454	0 <sup>b</sup>	N/A	0
Common from control cohort	266	3 (1.1%) <sup>b</sup>	N/A	MIR-526B

Notes: Significant difference was found between data a and data b (P < 0.05) (a two sided Fisher's exact test). N/A: not available.

ID (*MAPT* [30], *MEF2C* [31,32], *PAK2* [33]). Copy number changes of these targets could affect their regulation by miRNA. Target genes for CNV miRNAs detected in subjects with autism were also found to be enriched in MAPK pathway [21].

For familial CNVs we noted that although the number of miRNAs is much lower than in *de novo* or DECIPHER

pathogenic CNV groups, and not different than in common CNVs, the number of miRNA target genes was significantly higher than in common CNVs. This might suggest potential phenotypic impact of these familial CNVs. In our cohort we had two familial CNVs, predisposing to ID: 16p11.2 and 1q21.1. These CNVs are well-known for their heterogeneous phenotypes and familial or

*de novo* occurrence. The 1q21.1 CNV contains a single miRNA of unknown role (miR-5087), while the 16p11.2 paternal duplication covers 2 miRNAs: miR-3680-3p and miR-3680-5p and the *de novo* 16p11.2 deletion has no miRNA content (Additional file 1: Table S2). Consequences of the variability in copy number and function of miRNA integral to these CNVs is yet unknown. Understanding the role of miRNAs and miRNA targets in familial CNVs is of interest since this type of CNV represents a significant and clinically relevant interpretational challenge that could be guided by their miRNA features.

The ultimate proof that miRNAs influence the pathogenicity of genomic changes will come from an empirical confirmation of their copy number and expression change, and influence on the expression of their targets. The limitation of our own and other studies is due to their dependence on accurate prediction of miRNA targets and the fact that miRNA numbers increase dramatically between different versions of miRBase database. Therefore, comparison between different studies in terms of miRNA content in CNVs is challenging. Furthermore, some miRNAs may still represent false positive discoveries [10]. Unfortunately, we do not have miRNA expression data from our CNVs as RNA was not available. However, a recent study by Garcia-Orti et al. [34], demonstrated 10% of 259 studied miRNAs from regions of gain and loss detected in AML had significant change in expression concordant with the type of copy number change.

A more recently published study highlighted the significance of specific CNV-miRNAs and their targets in autism [21]. It assessed the content and function of 378 autism-associated CNVs, and detected 71 miRNAs. Five miRNAs were previously reported in ASD and 3 were known to have neuronal function. In our study, among the 84 miRNAs in the *de novo* CNVs, 3 were found to be associated with neurodegenerative disorders (miRNA-7, miRNA-9, miRNA-106b) [35-37] and 1 with ID (miRNA-9) [37].

Our analysis of miRNAs and miRNA targets related to CNVs is a first attempt to evaluate their role in a patient cohort manifesting idiopathic intellectual disabilities. Additional studies of cohorts of subjects with ID would benefit further evaluation of the apparent increase in miRNA content in *de novo* CNVs demonstrated in this study. In clinical practice, the interpretation of miRNAs that occur in patient CNVs is frequently challenging, particularly if they are the only genes that are included in the CNV. With global investigations of miRNAs in subjects with ID, followed by their expression analysis, our understanding of the role of miRNA in ID pathogenesis will be further improved.

## Conclusions

Our findings support a possible role of miRNA copy number change in cognition and/or CNV-mediated

developmental delay based on increased number of miRNA and miRNA target genes in *de novo* versus common CNVs in subjects with ID as well as their expression profile and participation in pathways. Systematic analysis of expression/function of miRNAs in addition to coding genes integral to CNVs could uncover new causes of ID.

## Methods

### Subjects

213 subjects with idiopathic ID were recruited for array CGH analysis by clinical geneticists across Canada. Selected individual or groups of cases were reported previously [38-42]. A cohort of 32 cognitively normal subjects had array testing as internal control samples. All of the subjects were tested by either Agilent 105 K Oligo array (227 subjects) or NimbleGen array (18 subjects). Forty syndromic genomic regions were selected from the DECIPHER database (<http://decipher.sanger.ac.uk/>) [43] and represented patients with neurodevelopmental delay associated with CNV findings of established potential for pathogenicity. The use of the DNA from these patients in our cohort was approved by the Committee for Ethical Review of Research involving Human Subjects, University of British Columbia. All subjects gave written informed consent for participation in the study and anonymized data were used for CNV/miRNA analysis.

### Array CGH

Genomic DNA was extracted from peripheral blood using PUREGENE DNA Isolation Kits (Gentra, Minneapolis, MN). A pool of normal male or female control DNAs (Promega, Madison, WI) was used as reference DNA matching the sex of the proband samples.

Agilent 105 K array analysis was performed according to the protocol provided by the company (version 4.0, June 2006, Agilent Technologies, CA, USA) [44]. Feature Extraction software (version 8.1.1.1, Agilent Technologies) rendered image analysis using the manufacturer's recommended settings (CGH\_v4\_95) and human genome assembly hg18. The minimum absolute average of log<sub>2</sub> ratio was 0.25.

Higher-resolution 385 K oligonucleotide genome array CGH was performed courtesy of NimbleGen. Array log<sub>2</sub> ratio > ±0.2 was used for a segmentation (region).

For both the Agilent and Nimblegen array platforms, ≥3 consecutive probes were required for a significant CNV call. CNVs that overlapped in genomic coverage were considered to represent the same CNV loci.

### Types of CNVs

All detected CNVs were grouped into 3 subgroups (*de novo*, familial and common CNVs) based on the criteria

described previously [45]. Briefly, CNVs completely overlapping with variants reported in at least two studies in the Database of Genomic Variants (DGV) (<http://projects.tcag.ca/variation/>) or in our internal controls were considered common CNVs; CNVs that overlapped partially (<50%) or did not overlap with CNVs reported in the DGV or our internal controls were called rare CNVs including de novo (not detected in proband's parents) and familial CNVs (inherited from either of parents). All unique CNVs (*de novo* and familial) were confirmed by an independent secondary method (such as FISH method) and only single copy gains or losses were identified.

### Bioinformatics analysis

The complete list of miRNAs in the whole genome was downloaded from miRBase v19 (<http://www.mirbase.org/>) [46], containing >2000 mature miRNAs in human. The miRNAs in different CNV subgroups were obtained by using intersecting tool in Galaxy (<https://main.g2.bx.psu.edu/>) [47].

The expression profiles of miRNAs were obtained by using mimiRNA database (<http://mimirna.centenary.org.au/mep/formulaire.html>) [24] and microRNA.org - Targets and Expression (<http://www.microrna.org/microrna/home.do>) [25]. Both web-tools provide experimentally derived miRNA expression data after inputting a specific miRNA name.

The genomic locations of 19,905 human protein coding genes (PCG) were extracted from the MISO database (<http://genes.mit.edu/burgelab/miso/index.html>). Gene content of different CNV subgroups was obtained by considering genes within or covering CNV regions.

WebGestalt2 (WEB-based GENE SeT Analysis Toolkit V2) (<http://bioinfo.vanderbilt.edu/webgestalt/>) is a publicly available web-tool for functional enrichment analysis of gene sets using a web-based integrated data mining system [48]. Using hypergeometric test, the top 10 groups of protein coding genes from each CNV subgroup that are targets for miRNA were generated by this tool and ranked by adjP (p value adjusted by multiple test adjustment). Among these miRNA target genes, only those with adjP < 0.05, i.e. distributed in CNVs more often than expected when compared to reference genome, were selected for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. Similarly, the top ten KEGG pathways enriched for miRNA target gene were generated by the same tool for each CNV subgroup. Only the top ten enriched pathways with adjP < 0.05 were selected and compared between different CNV subgroups. The WebGestalt2 tool was also used to identify miRNAs that target the target genes for each CNV group.

### Statistical analysis

#### All statistical analyses were performed in R 2.12

*MiRNA and protein coding gene (PCG) coverage relative to random distribution:* for each CNV/CNVR subgroup we generated a set of random CNV/CNVRs with similar length distribution and total genome coverage. The number of miRNA regions and PCG regions affected by these random CNV/CNVRs was assessed. This operation was repeated 1000 times for each of the 5 CNV subgroups, generating a series of values that served as a measurement of the miRNA and PCG coverage of CNV/CNVRs under a random distribution. We performed a two-tailed test to compute the p value of the actual miRNA and PCG coverage of the 5 CNV/CNVR subsets. The p values were calculated as the number of values falling above (or below) the observed miRNA and gene coverage of the actual CNVRs datasets.

*Comparison of miRNA and PCG content:* we computed the number of miRNA and PCG in the individual CNVs relative to their size i.e. the number of miRNA and PCG per Mb. We compared the subgroups to each other using a Wilcoxon signed-rank test to assess whether a group has a significantly different miRNA and PCG density.

In order to determine the significance of the number miRNA related to brain function, we compared the fraction of brain-function related miRNA in the pool of available CNV-miRNA between each pair of CNV subgroups. The statistical p value was computed using a two-sided Fisher's exact test. Similarly, we performed two-sided Fisher's exact tests to compare the fraction of genes targeted by miRNA in the different CNV subgroups using the total number of genes encompassed by the CNVs.

### Additional file

**Additional file 1: Genomic coordinates of CNV/CNVRs and miRNAs used in this study.**

### Abbreviations

CNVs: Copy number variations; CNVRs: Copy number variant regions; DECIPHER: Database of chromosomal imbalance and phenotype in humans using ensembl resources; ID: Intellectual disability; KEGG: Kyoto encyclopedia of genes and genomes; MAPK: Mitogen-activated protein kinase; MAPK9: Mitogen-activated protein kinase 9; MAPT: Microtubule-associated protein tau; MEF2C: Myocyte enhancer factor 2C; miRNA: MicroRNA; MKNK2: MAP kinase interacting serine/threonine kinase 2; PAK2: P21 protein (Cdc42/Rac)-activated kinase 2; PCG: Protein coding gene; pri-miRNAs: Primary transcript miRNA; RPS6KA1: Ribosomal protein S6 kinase; 90 kDa: Polypeptide 1; SNPs: Single nucleotide polymorphisms; TAOK2: TAO kinase 2; WebGestalt2: WEB-based GENE SeT Analysis Toolkit V2.

### Competing interests

We declare no conflict of interest in our manuscript titled as "miRNA and miRNA target genes in copy number variations occurring in individuals with intellectual disability".

### Authors' contributions

YQ performed data acquisition, data analysis, and drafting of manuscript; CB initiated the project and analyzed the data; EM performed statistical and



bioinformatics analyses; SL recruited clinical cases and reviewed the manuscript; PP supervised statistical and bioinformatics analyses and reviewed the manuscript; ES supervised and designed the study, helped with data interpretation, and critically revised the manuscript. All authors read and approved the final manuscript.

#### Acknowledgements

This work was supported by funding from the Canadian Institutes for Health Research (CIHR) (MOP 74502; PI: ERS), Canadian Foundation for Innovation-Leading Edge Fund (CFI-LEF; PI: MESL) and the BC Knowledge Development Fund (MOP 64217; PI: MESL), and Establishment Grant funding from the Michael Smith Foundation for Health Research. PP was supported by a career award from the Michael Smith Foundation for Health Research, a CIHR New Investigator award, the Canadian Foundation for Innovation, and the National Institutes of Health (GM076990). MESL and ERS are Career Scholars supported by the Michael Smith Foundation for Health Research. The authors appreciate the collaboration and support of the participating subjects and their families. This study makes use of data generated by the DECIPHER Consortium. A full list of centres who contributed to the generation of the data is available from <http://decipher.sanger.ac.uk> and via email from [decipher@sanger.ac.uk](mailto:decipher@sanger.ac.uk). Funding for the project was provided by the Wellcome Trust.

#### Author details

<sup>1</sup>Department of Pathology and Lab Medicine, BC Child and Family Research Institute, University of British Columbia (UBC), Vancouver, BC V5Z 4H4, Canada. <sup>2</sup>Department of Medical Genetics, BC Child and Family Research Institute, University of British Columbia (UBC), Vancouver, BC V6H 3N1, Canada. <sup>3</sup>Centre for High-throughput Biology, 177 Michael Smith Laboratories, UBC, Vancouver, BC V6T 1Z4, Canada.

Received: 23 March 2013 Accepted: 6 August 2013

Published: 10 August 2013

#### References

- Lee RC, Feinbaum RL, Ambros V: The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 1993, **75**:843–854.
- Wightman B, Ha I, Ruvkun G: Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 1993, **75**:855–862.
- Kim VN: MicroRNA biogenesis: coordinated cropping and dicing. *Nat Rev Mol Cell Biol* 2005, **6**:376–385.
- Lewis BP, Burge CB, Bartel DP: Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005, **120**:15–20.
- Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, Lim B, Rigoutsos I: A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* 2006, **126**:1203–1217.
- Vasudevan S: Posttranscriptional upregulation by microRNAs. *Wiley Interdiscip Reviews* 2011, **3**:311–330.
- Sethupathy P, Collins FS: MicroRNA target site polymorphisms and human disease. *Trends Genet* 2008, **24**:489–497.
- Duan S, Mi S, Zhang W, Dolan ME: Comprehensive analysis of the impact of SNPs and CNVs on human microRNAs and their regulatory genes. *RNA Biol* 2009, **6**:412–425.
- Felekis K, Voskarides K, Dweep H, Sticht C, Gretz N, Deltas C: Increased number of microRNA target sites in genes encoded in CNV regions. Evidence for an evolutionary genomic interaction. *Mol Biol Evol* 2011, **28**:2421–2424.
- Marcinkowska M, Szymanski M, Krzyzosiak WJ, Kozłowski P: Copy number variation of microRNA genes in the human genome. *BMC Genomics* 2011, **12**:183.
- Schiffman JD, Lorimer PD, Rodic V, Jahromi MS, Downie JM, Bayerl MG, Sanmann JN, Althof PA, Sanger WG, Barnette P, et al: Genome wide copy number analysis of paediatric Burkitt lymphoma using formalin-fixed tissues reveals a subset with gain of chromosome 13q and corresponding miRNA over expression. *Brit J Haematol* 2011, **155**:477–486.
- Shim SM, Nam HY, Lee JE, Kim JW, Han BG, Jeon JP: MicroRNAs in human lymphoblastoid cell lines. *Crit Rev Eukar Gene* 2012, **22**:189–196.
- Bandiera S, Hatem E, Lyonnet S, Henrion-Caude A: microRNAs in diseases: from candidate to modifier genes. *Clinical genetics* 2010, **77**:306–313.
- Esteller M: Non-coding RNAs in human disease. *Nature reviews* 2011, **12**:861–874.
- Ghahramani Seno MM, Hu P, Gwadry FG, Pinto D, Marshall CR, Casallo G, Scherer SW: Gene and miRNA expression profiles in autism spectrum disorders. *Brain research* 2011, **1380**:85–97.
- Provost P: MicroRNAs as a molecular basis for mental retardation, Alzheimer's and prion diseases. *Brain research* 2010, **1338**:58–66.
- Talebizadeh Z, Butler MG, Theodoro MF: Feasibility and relevance of examining lymphoblastoid cell lines to study role of microRNAs in autism. *Autism Res* 2008, **1**:240–250.
- Bian S, Sun T: Functions of noncoding RNAs in neural development and neurological diseases. *Mol Neurobiol* 2011, **44**:359–373.
- Willemsen MH, Valles A, Kirkels LA, Mastebroek M, Olde Loohuis N, Kos A, Wissink-Lindhout WM, de Brouwer AP, Nillesen WM, Pfundt R, et al: Chromosome 1p21.3 microdeletions comprising DPYD and MIR137 are associated with intellectual disability. *J Med Genet* 2011, **48**:810–818.
- de Pontual L, Yao E, Callier P, Faivre L, Drouin V, Cariou S, Van Haeringen A, Genevieve D, Goldenberg A, Oufadem M, et al: Germline deletion of the miR-17 approximately 92 cluster causes skeletal and growth defects in humans. *Nature genetics* 2011, **43**:1026–1030.
- Vaishnavi V, Manikandan M, Tiwary BK, Munirajan AK: Insights on the functional impact of MicroRNAs present in autism-associated copy number variants. *PLoS one* 2013, **8**:e56781.
- Kosik KS: The neuronal microRNA system. *Nat Rev Neurosci* 2006, **7**:911–920.
- Szulwach KE, Jin P, Alisch RS: Noncoding RNAs in mental retardation. *Clinical genetics* 2009, **75**:209–219.
- Ritchie W, Flamant S, Rasko JE: mimiRNA: a microRNA expression profiler and classification resource designed to identify functional correlations between microRNAs and their targets. *Bioinformatics (Oxford, England)* 2010, **26**:223–227.
- Betel D, Koppal A, Agius P, Sander C, Leslie C: Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome biology* 2010, **11**:R90.
- Wu X, Zhang D, Li G: Insights into the regulation of human CNV-miRNAs from the view of their target genes. *BMC Genomics* 2012, **13**:707.
- Aoki Y, Niihori T, Narumi Y, Kure S, Matsubara Y: The RAS/MAPK syndromes: novel roles of the RAS pathway in human genetic disorders. *Hum Mutat* 2008, **29**:992–1006.
- Houge G, Rasmussen IH, Hovland R: Loss-of-function CNKSR2 mutation is a likely cause of non-syndromic X-linked intellectual disability. *Mol Syndromol* 2012, **2**:60–63.
- McMillan EL, Kamps AL, Lake SS, Svendsen CN, Bhattacharyya A: Gene expression changes in the MAPK pathway in both Fragile X and down syndrome human neural progenitor cells. *Am J Stem Cells* 2012, **1**:154–162.
- Shaw-Smith C, Pittman AM, Willatt L, Martin H, Rickman L, Gribble S, Curley R, Cumming S, Dunn C, Kalaitzopoulos D, et al: Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. *Nat Genet* 2006, **38**:1032–1037.
- Nowakowska BA, Obersztyn E, Szymanska K, Bekiesinska-Figatowska M, Xia Z, Ricks CB, Bocian E, Stockton DW, Szczaluba K, Nawara M, et al: Severe mental retardation, seizures, and hypotonia due to deletions of MEF2C. *Am J Med Genet B Neuropsychiatr Genet* 2010, **153B**:1042–1051.
- Zweier M, Gregor A, Zweier C, Engels H, Sticht H, Wohlleber E, Bijlsma EK, Holder SE, Zenker M, Rossier E, et al: Mutations in MEF2C from the 5q14.3q15 microdeletion syndrome region are a frequent cause of severe mental retardation and diminish MECP2 and CDKL5 expression. *Hum Mutat* 2010, **31**:722–733.
- Quintero-Rivera F, Sharifi-Hannauer P, Martinez-Agosto JA: Autistic and psychiatric findings associated with the 3q29 microdeletion syndrome: case report and review. *Am J Med Genet* 2010, **152A**:2459–2467.
- Garcia-Orti L, Cristobal I, Cirauqui C, Guruceaga E, Marcotegui N, Calasanz MJ, Castello-Cros R, Odero MD: Integration of SNP and mRNA arrays with microRNA profiling reveals that MiR-370 is upregulated and targets NF1 in acute myeloid leukemia. *PLoS One* 2012, **7**:e47717.
- Doxakis E: Post-transcriptional regulation of alpha-synuclein expression by mir-7 and mir-153. *J Biol Chem* 2010, **285**:12726–12734.
- Wang H, Liu J, Zong Y, Xu Y, Deng W, Zhu H, Liu Y, Ma C, Huang L, Zhang L, et al: miR-106b aberrantly expressed in a double transgenic mouse model for Alzheimer's disease targets TGF-beta type II receptor. *Brain Res* 2010, **1357**:166–174.

37. Xu XL, Zong R, Li Z, Biswas MH, Fang Z, Nelson DL, Gao FB: **FXR1P but not FMRP regulates the levels of mammalian brain-specific microRNA-9 and microRNA-124.** *J Neurosci* 2011, **31**:13705–13709.
38. Gibson WT, Harvard C, Qiao Y, Somerville MJ, Lewis ME, Rajcan-Separovic E: **Phenotype-genotype characterization of alpha-thalassemia mental retardation syndrome due to isolated monosomy of 16p13.3.** *Am J Med Genet* 2008, **146A**:225–232.
39. Harvard C, Malenfant P, Koochek M, Creighton S, Mickelson EC, Holden JJ, Lewis ME, Rajcan-Separovic E: **A variant Cri du Chat phenotype and autism spectrum disorder in a subject with de novo cryptic microdeletions involving 5p15.2 and 3p24.3–25 detected using whole genomic array CGH.** *Clin Genet* 2005, **67**:341–351.
40. Tyson C, Harvard C, Locker R, Friedman JM, Langlois S, Lewis ME, Van Allen M, Somerville M, Arbour L, Clarke L, et al: **Submicroscopic deletions and duplications in individuals with intellectual disability detected by array-CGH.** *Am J Med Genet* 2005, **139**:173–185.
41. Tyson C, McGillivray B, Chijiwa C, Rajcan-Separovic E: **Elucidation of a cryptic interstitial 7q31.3 deletion in a patient with a language disorder and mild mental retardation by array-CGH.** *Am J Med Genet* 2004, **129A**:254–260.
42. Rajcan-Separovic E, Harvard C, Liu X, McGillivray B, Hall JG, Qiao Y, Hurlburt J, Hildebrand J, Mickelson EC, Holden JJ, et al: **Clinical and molecular cytogenetic characterisation of a newly recognised microdeletion syndrome involving 2p15–16.1.** *J Med Genet* 2007, **44**:269–276.
43. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, Van Vooren S, Moreau Y, Pettett RM, Carter NP: **DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources.** *Am J Hum Genet* 2009, **84**:524–533.
44. Fan YS, Jayakar P, Zhu H, Barbouth D, Sacharow S, Morales A, Carver V, Benke P, Mundy P, Elsas LJ: **Detection of pathogenic gene copy number variations in patients with mental retardation by genomewide oligonucleotide array comparative genomic hybridization.** *Hum Mutat* 2007, **28**:1124–1132.
45. Qiao Y, Tyson C, Hrynchak M, Lopez-Rangel E, Hildebrand J, Martell S, Fawcett C, Kasmara L, Calli K, Harvard C, et al: **Clinical application of 2.7M cytogenetics array for CNV detection in subjects with idiopathic autism and/or intellectual disability.** *Clin Genet* 2013, **83**:145–154.
46. Kozomara A, Griffiths-Jones S: **miRBase: integrating microRNA annotation and deep-sequencing data.** *Nucleic acids research* 2011, **39**:D152–157.
47. Goecks J, Nekrutenko A, Taylor J: **Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.** *Genome Biol* 2010, **11**:R86.
48. Zhang B, Kirov S, Snoddy J: **WebGestalt: an integrated system for exploring gene sets in various biological contexts.** *Nucleic Acids Res* 2005, **33**:W741–748.

doi:10.1186/1471-2164-14-544

**Cite this article as:** Qiao et al.: miRNA and miRNA target genes in copy number variations occurring in individuals with intellectual disability. *BMC Genomics* 2013 **14**:544.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

