The impact of chromosomal rearrangements on regulation of gene expression

Louise Harewood* and Peter Fraser

Nuclear Dynamics Programme, The Babraham Institute, Babraham Research Campus, Cambridge CB22 3AT, UK

Received March 27, 2014; Revised May 20, 2014; Accepted June 4, 2014

The effects that coding region single-nucleotide polymorphisms or mutations have on gene expression have been well documented, predominantly owing to their association with disease. The effects of structural chromosomal rearrangements are also receiving increasing attention with the development of new techniques that allow accurate, high-resolution data, whether genomic interaction or transcriptome data, to be generated right down to the single-cell level. Over the past 18 months, these advances in experimental techniques have been used to further confirm and delineate the substantial effects that chromosome rearrangements can have on the regulation of gene expression and provide evidence of direct links between the two.

INTRODUCTION

Chromosome rearrangements occur in two broad forms—balanced and unbalanced. Balanced, or copy number neutral, refers to those in which there is no net gain or loss of genetic material, though some rearrangement has occurred—this includes reciprocal translocations and inversions. Unbalanced rearrangements are those in which genetic material is either gained or lost and can include copy number variant (CNV) regions, caused by deletions or duplications, or the inheritance of unbalanced translocation products (Figs. 1 and 2). Both types of rearrangement have been shown to have impacts on gene expression through a variety of different mechanisms, some of which are outlined in this review.

CHROMOSOME REARRANGEMENT BREAKPOINTS CAN DIRECTLY DISRUPT GENE EXPRESSION

The association of balanced chromosomal rearrangements, such as a reciprocal translocations or inversions, with aberrant gene expression levels may have many underlying causes, the simplest of which is direct disruption of one or more genes or regulatory elements by one or more of the rearrangement breakpoints.

Balanced chromosomal rearrangements have long since been associated with both gene expression changes and clinical phenotypes, with causative loci/genes for many Mendelian disorders being discovered via mapping of disease-associated rearrangement breakpoints [reviewed in (1)]. Rearrangements can also result in the production of functional fusion genes, placing part of one gene under control of regulatory elements for another or fusing two gene coding regions. This is predominantly seen in cancer, with the Philadelphia chromosome formed by the t(9;22)(q34;q11) translocation in chronic myelogenous leukaemia causing ABL1 to be placed under the control of the BCR gene, probably being the most famous example (2,3). Fusion genes can also be found in constitutional disorders, such as aromatase excess syndrome (OMIM 139300), where an inversion places the aromatase-encoding CYP19 gene under control of a ubiquitously expressed cryptic promoter, causing overexpression of aromatase and the associated phenotype (4,5).

POSITION EFFECT IMPACTS ON GENE EXPRESSION REGULATION

This formation of cryptic promoters is one way in which gene expression can be affected by chromosomal rearrangements. Similarly, whilst breakpoints may fall within the coding region of a gene and therefore cause perturbations in expression, they have also been shown to occur outside of the gene itself but affect regulation by causing the disruption of cis-regulatory elements. This phenomenon, known as position effect or disruption, has been reported for a number of human disorders and also associated mouse models (6–8), with breakpoints occurring both proximally and distally to the affected genes and at distances up to 1.5 Mb (9). Whilst this points to regulatory elements being positioned some distance away from the gene

*To whom correspondence should be addressed. Tel: +44 1223496000; Email: lajharewood@gmail.com

© The Author 2014. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com
that they control, most \textit{cis}-interactions occur in regions nearer to the gene in question, with studies showing an inverse correlation between interaction frequency and distance from the promoter (10–12). A recent study combined oligonucleotide capture technology, 3C and high-throughput sequencing into a new method, dubbed Capture-C, and used it to examine hundreds of \textit{cis}-interactions in mice in a single experiment (11). They showed that most \textit{cis}-interactions occurred in a 600-kb window around the promoter but also reiterated the previous finding of regulatory elements sometimes being separated from the gene they control not only by distance but also by coding regions of other genes (6,11,13). This potentially explains the effects that chromosome rearrangements can have on the expression of genes not only near to, but also at some distance from, the breakpoints.

Position effect can also be caused by genes being placed into anomalous chromatin environments. One example of this in a human disorder has recently been proposed, where a translocation between heterochromatic chromosome band 15p11.2 and euchromatic band 16q12.1 was associated with a neurological phenotype. Gene expression studies in this patient led the authors to suggest position effect, similar to position effect variegation seen in \textit{Drosophila} (14), had occurred where heterochromatic genes were enhanced and euchromatic genes silenced as a result of their juxtaposition to constitutive heterochromatin, in this case owing to the t(15;16)(p11.2;q12.1) translocation (15). A similar mechanism has also been proposed for a t(X;2)(q23;q33) associated with an Incontinentia Pigmenti (IP)-like phenotype in a female carrier, where some genes situated on the translocated region of the X chromosome (Xq24-qter) were suggested to be downregulated because of their proximity to heterochromatic band 2q34 (16). One of the downregulated genes in this translocated region was \textit{IKBKG}, the causative gene of IP, situated in band Xq28 (OMIM 300248), the partial silencing of which would presumably contribute to the phenotype of this patient.

This phenomenon has also been observed in cancer with one example involving an unbalanced translocation (+der(2)t(1;2)(q12;p13)) in B-cell lymphoma showing chromatin alterations along the derivative chromosome and abnormal heterochromatic structures, dubbed aberrant heterochromatin foci (aHCF), within the nucleus. Chromosome 2p sequences adjacent to the breakpoint showed repressive heterochromatin marks (H4K20me3, H3K9me3 and HP1), delayed replication and a number of deregulated genes. The authors proposed that chromosome band 1q12 rearrangements corresponded to ‘a new class of chromosome rearrangements with the power, via novel long range heterochromatin-dependent mechanisms, to profoundly perturb gene organization and function’ (17).

Another type of position effect that has the potential to affect gene expression and has been postulated to play a role in human disease is telomere position effect (TPE). This is a phenomenon by which genes situated near to telomeres can show changes in expression that are correlated (whether positively or negatively) with telomere length. A recent paper (18) has suggested that TPE
may contribute to the phenotype of the progressive muscle wasting disease, facioscapulohumeral muscular dystrophy (FSHD1, OMIM 158900). Subclones of cells with different telomere lengths were artificially generated, and expression of the DUX4 gene analysed. This gene is located within D4Z4 repeats found in the subtelomeric region of chromosome 4q. Contraction or deletion of these repeats has been associated with, though is not the sole basis of, the FSHD phenotype. In 2010, a ‘Unifying Genetic Model’ for FSHD was proposed in which contraction of the D4Z4 repeats was accompanied by a toxic gain of function of stable DUX4 transcripts (19), and the recent DUX4 TPE paper showed that DUX4 expression was upregulated over 10-fold in FSHD cells with short telomeres, with expression progressively increasing as telomere length shortened, suggesting that TPE may play a role in the age-related phenotype of FSHD patients (18). FSHD is known to be caused by deletions of 4q35 that cause shortening of the D4Z4 repeat region or via translocations between the D4Z4 repeats on chromosome 4q and highly homologous sequences located in the subtelomeric region of chromosome 10q (20).

**Figure 2.** Examples of chromosome rearrangements affecting gene expression. Top left: normal DNA interactions can be disrupted by translocations causing derivative chromosomes to alter their location within the nucleus. A translocation causes abrogation of normal long-range cis-interactions on the red chromosome (visible on the left panel) and causes new interactions to occur between regions of the different chromosomes (right). Top right: an inversion with breakpoints (red dashed lines) situated within two genes in opposite orientations causes both to be placed under the control of cryptic promoters and fusion transcripts to be produced. Middle left: an intergenic deletion that does not involve coding regions of genes but deletes regulatory elements. In the normal situation (top), gene A is repressed and B expressed. Deletion of the silencer of A and the enhancer of B causes a change in the expression of both genes. Middle right: duplications can result in an increase in the copy number of genes and generally an associated increase in their expression. Bottom: chromosome rearrangements may disrupt loops that occur between genes in multigene complexes, causing hierarchical gene expression effects. In this example, the association with gene 1 at a site of transcription causes gene 2 to associate, which in turn causes gene 3 to do the same. Disruption of gene 2, preventing this association from occurring, also prevents gene 3 from initiating, meaning that a chromosome rearrangement involving gene 2 affects not only the expression of that gene but also that of others downstream in the cascade.
The combination of these chromosomal rearrangements causing shortening of the D4Z4 repeats and the TPE produces alterations in expression of DUX4, potentially forming the basis of the FSHD phenotype.

As well as effects on expression of genes in cis, trans-effects (on chromosomes other than those involved in the rearrangements) can also be observed as a result of chromosomal rearrangements and these effects can be genome-wide.

CHROMOSOME REARRANGEMENT-INDUCED NUCLEAR REORGANIZATION CAN EXERT GENOME-WIDE EFFECTS ON EXPRESSION

Within the nucleus, chromosomes are organized into distinct territories that occupy non-random, cell-type-specific positions (21,22). The highly conserved arrangement of genetic material suggests that perturbations in this may have a sizeable effect on gene expression. In the case of the aforementioned translocation in B-cell lymphoma, as well as alterations in chromatin marks and formation of aHCF, the derivative chromosome was also shown to occupy a more peripheral position within the nucleus than the normal chromosome 2 (17). This aberrant nuclear positioning of chromosome territories as a result of chromosome translocations has been previously reported (21,23–25) and of note is the finding that whilst repositioning of the derivative chromosomes can have an effect on the genes situated both near, and distant, to the breakpoints, these changes are not restricted to only those chromosomes involved in the rearrangement. Transcriptome analyses of balanced constitutional t(11;22)(q23;q11) carriers as compared with cytogenetically normal controls showed a number of differentially expressed transcripts that were not restricted to the chromosomes involved in the translocation but situated on nearly every chromosome throughout the genome (26).

These genome-wide gene expression changes may be due to the fact that chromosomal rearrangements can impact on the spatial organization of chromosomes and cause a general nuclear reorganization. For example, carriers of the constitutional t(11;22)(q23;q11) exhibit anomalous nuclear positioning of not only chromosomes 11 and 22 but also chromosome 17, a gene-dense chromosome that was shifted to a significantly more peripheral position (23). This suggests that the relocation of derivative chromosomes has a knock on effect that results in displacement of other chromosomes from their usual nuclear positions. This repositioning of derivative, and other, chromosomes within the nucleus and associated changes in gene expression may have many causes. Although unequivocal links between nuclear location and gene expression have yet to be ascertained, there are multiple lines of evidence to suggest that expression of a gene is dependent not only on regulatory elements and local chromatin context but also on the position of that gene within the nucleus [reviewed in (27–30)]. The arrangement of the chromosome territories and the finding that active and inactive copies of monoaellically expressed genes occupy different positions within the nucleus (31) suggests that nuclear position and gene expression are directly linked. However, the view that increases in gene expression correlate with a shift towards the centre of the nucleus, and vice versa, (32–34) is yet to be conclusively proven. The finding that only a limited number of genes on repositioned chromosomes in t(11;22) cells were detectably misexpressed was consistent with previous studies in which chromosome position was experimentally manipulated (32,33), suggesting that nuclear environment is not a contributing factor in the control of expression for every gene within the genome.

Reorganization of chromosomes within the nucleus can therefore result in disruption of both intrachromosomal (cis) and interchromosomal (trans) interactions. A study examining cis-contacts between regions in the human genome has shown that they occur with a probability that is inversely correlated with their genomic separation distance, although very long-range cis-interactions have been robustly demonstrated (12). Co-associations of distant genes have also been shown to occur at ‘transcription factories’ (35–39)—discrete foci containing high concentrations of active RNA Polymerase II at which actively transcribed alleles have been shown to be positioned — or around aggregations of splicing-related factors, termed ‘Nuclear Speckles’ (40,41). Therefore, relocation of not only the derivative chromosomes but also other chromosomes within the nuclei of translocation carriers may cause disruption of these normal interactions and consequently have an effect on gene expression.

Trans-interactions can also be affected by chromosome rearrangements. Although each chromosome occupies a specific territory within the nuclear space, the boundaries between these are not impenetrable and intermingling between chromosomes can occur. This mixing of territories is thought to be transcription dependent as transcriptional inhibition affects the amount of interweaving observed (42–44). As well as long-range cis-interactions, interactions between genes from different chromosomes have also been observed at transcription factories (35–39). If these transcription-dependent associations between chromosomes are enough to influence their arrangement within the nuclear space, it is not unreasonable to hypothesize that an alteration in nuclear organization, caused by a chromosome rearrangement, could have an effect on transcription and therefore gene expression. In support of this is a 2013 study in which a single-cell strategy was devised using TALE nucleases to specifically disrupt sites within gene loops that are known to engage in chromosomal contact within a multigene complex (45). By doing this, they demonstrated that this disruption produces hierarchical gene expression effects on interacting genes, suggesting that contact between these is required for normal transcription. Disruption of these contacts via chromosomal rearrangements may, therefore, produce a cascade of alterations on gene expression.

Gene expression and nuclear organization appear to be inextricably linked but so far, direct evidence linking nuclear repositioning with alterations in expression has been somewhat speculative, with proof of direct disruption of specific genic interactions as a result of nuclear relocation remaining elusive. Further studies such as those using 4C-Seq, or other chromosome conformation capture (3C)-based techniques, to compare interactions observed in cells with chromosome rearrangements as compared with cytogenetically normal individuals, or even between affected and unaffected alleles in the same cell, will hopefully shed more light on just how far the effects of chromosome rearrangements reach with regards to gene expression and regulation.

A different approach to examine the changes in gene expression between normal and rearranged chromosomes was recently...
GAINS/LOSSES OF DNA SEGMENTS HAVE FAR-REACHING EFFECTS ON GENE EXPRESSION

In addition to the rearrangement of genetic material caused by chromosomal abnormalities such as translocations and inversions, gains or losses of genetic material, i.e. deletions and duplications, have long since been shown to have an effect on gene expression. Large cytogenetically visible rearrangements have been associated with numerous different syndromes, and the advent of technologies, such as comparative genomic hybridization (CGH) in the early 1990s (47) and later array-based CGH (48,49), allowed the discovery of further cytogenetically cryptic deletions and duplications that were both associated with specific phenotypes (50) and also present within normal populations (51,52).

A study from 2007 examined the impact of CNVs on gene expression phenotypes and showed that CNVs accounted for 17.7% of gene expression variation (53). Since then, the resolution of CNV detection has improved, resulting in estimates of the number of CNV regions per genome being dramatically increased (54,55), thereby presumably also increasing the estimation of their overall effects on gene expression. The effects that CNVs exert on expression have been shown to vary through developmental time points and also appear to show cell-type specificity (56,57).

Most genes situated within CNV regions show expression levels that are positively correlated with gene copy number, i.e. genes within duplicated regions that show an increased copy number also have increased expression and the opposite is true for deletions. However, the expression levels of some genes show an inverse correlation (53,57–59). Many mechanisms have been proposed to explain this, from inaccurate breakpoint prediction, to immediate early genes controlling negative feedback loops (1), to steric hindrance of extra copies of gene products (60)—but precise mechanisms are yet to be elucidated. There are also genes that do not show significant expression changes, despite being either increased or reduced in copy number, suggesting that dosage compensation or buffering mechanisms may be in operation (61).

Genes that are situated near to rearrangement breakpoints but present in normal copy number can also show expression changes. For example, microdeletions of human chromosome 7q11.23 that cause Williams–Beuren syndrome (WBS, OMIM 194050) have been shown to have an effect on the expression of normal copy number genes situated both proximal and distal to deletion breakpoints, up to a distance of 6.5 Mb (62). A 2013 paper used circularized chromosome conformation capture sequencing (4C-Seq) (63) to examine genomic interactions around this microdeletion and determine whether this could explain expression changes seen in neighbouring, normal copy number genes (64). They showed that alterations of flanking genes may be attributed to abrogation of normal long-range interactions, with alterations in looping being observed between expression-affected neighbouring genes and the deleted region in WBS cells (64). Around the same time, a similar paper was published examining interactions disrupted in 22q11.2 deletion syndrome (OMIM 611867) by using the COMT gene, situated within the commonly deleted 3-Mb region, as a bait for a 4C-Seq experiment (65).

Alterations in expression of neighbouring, normal copy number genes have also been shown for other CNVs, including both deletions and duplications of the human 16p11.2 region (66) and mouse models of Smith–Magenis (deletion of chromosome 17p11.2 in human; OMIM 182290) and the reciprocal Potocki–Lupski (microduplication of 17p11.2 in human; OMIM 610883) syndromes, where gene expression changes were seen not only in the region flanking the rearrangement but along the whole length of the chromosome carrying the deletion/duplication (67). The expression changes seen in flanking genes are not necessarily correlated with the copy number change within the CNV segment, with duplications and deletions sometimes having similar effects. For example, genes mapping distally to the 16p11.2 deletion or duplication are shown to be similarly upregulated in both sets of carriers (66), showing that it is rearrangement of the interval that causes flanking effects and not the copy number change itself. Examples have also been shown of CNVs affecting the expression of genes situated throughout the genome (68). There are many possible explanations for these findings, including miRNAs being within the CNV region or disrupted by it (69), disruption of cis-regulatory elements or trans-interactions, downstream effects from disruption of regulatory or dosage-sensitive genes or modification of transcriptional control through chromatin structure alteration.

Deletions or duplications can therefore exert an effect on normal gene expression levels, not only by changing the number of copies of a gene and potentially its associated gene product but also by affecting other genes in the region of the breakpoints and further afield, and possibly also altering expression timing and location. These widespread, spatial and temporal effects all need to be considered when studying the contribution of CNVs to gene expression.

As well as classical CNVs, copy number changes in the genome can also occur through inheritance of unbalanced products from chromosome rearrangements, such as translocations. These can result in large regions of chromosomes being gained or lost. In a study of patients with Emanuel syndrome (OMIM 609029), a disorder caused by inheritance of an extra derivate chromosome 22 (+der(22)t(11;22)q23;q11) from a balanced t(11;22) translocation carrier, significant changes in gene expression were observed when comparing affected individuals with either cytogenetically normal controls or balanced t(11;22)
carriers. Consistently, a significant number of the differentially expressed transcripts mapped to chromosomes 11 and 22 and all genes mapping to trisomic regions showed an increase in expression. However, a large number of differentially expressed transcripts also mapped to other chromosomes, with every chromosome being represented, suggesting that the gain of large genomic regions can have many downstream or trans-effects on the expression of normal copy number genes (26).

CONCLUSIONS AND FUTURE PERSPECTIVES

New techniques are continually being designed and developed to allow both genomic interactions and gene expression to be more accurately interrogated. In late 2013, it was demonstrated that analysis of genomic interactions using HiC (12), an adapted version of 3C, is possible even within single cells (70), meaning that the averaging effects of these studies on population of cells can potentially be overcome. The same is true for gene expression analyses. Within the last 18 months, single-cell sequencing and transcriptome analyses have hit the limelight, with *Nature Methods* journal naming single-cell sequencing as its method of the year for 2013 (71). The data that will be generated from these sequencing experiments are likely to also kick-start the generation of a whole new wave of analysis tools and bioinformatic approaches that will enable genome organization and regulation to be investigated in much more depth than is currently possible. Hopefully, in the near future, we will be able to gain more insight into the effects that structural chromosome rearrangements, whether they be CNVs or balanced rearrangements, have on the regulation of gene expression, both locally and genome-wide.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the Biotechnology and Biological Science Research Council, UK.

REFERENCES


