

MINIREVIEW

# High-throughput analysis of chromosome translocations and other genome rearrangements in epithelial cancers

Scott Newman\* and Paul AW Edwards\*

## Abstract

Genes that are broken or fused by structural changes to the genome are an important class of mutation in the leukemias and sarcomas but have been largely overlooked in the common epithelial cancers. Large-scale sequencing is changing our perceptions of the cancer genome, and it is now being applied to structural changes, using the 'paired end' strategy. This reveals more clearly than before the extent to which many cancer genomes are rearranged and how much these rearrangements contribute to the mutational burden of epithelial tumors. In particular, there are probably many fusion genes, analogous to those found in leukemias, to be found in common cancers, such as breast carcinoma, and some of these will prove to be important in cancer diagnosis and treatment.

## Introduction

Somatic structural variations in the genome - referred to by cytogeneticists as translocations, inversions, duplications and insertions - can be powerful events in tumor evolution because they can create fusion genes. Fusion genes are formed when part of one gene is juxtaposed to another by a structural rearrangement, creating a hybrid transcript, or sometimes simply inserting a novel promoter upstream of a gene. These can be very powerful oncogenic mutations, not only increasing expression of a protein but also changing its activity, subcellular localization or binding specificity [1,2]. Such fusion genes are also clinically important, because some can predict outcome and determine management, and some may be targets for therapy [1]. For example, the *BCR-ABL* fusion gene defines a group of leukemias and is the target of treatment with the kinase inhibitor Glivec.

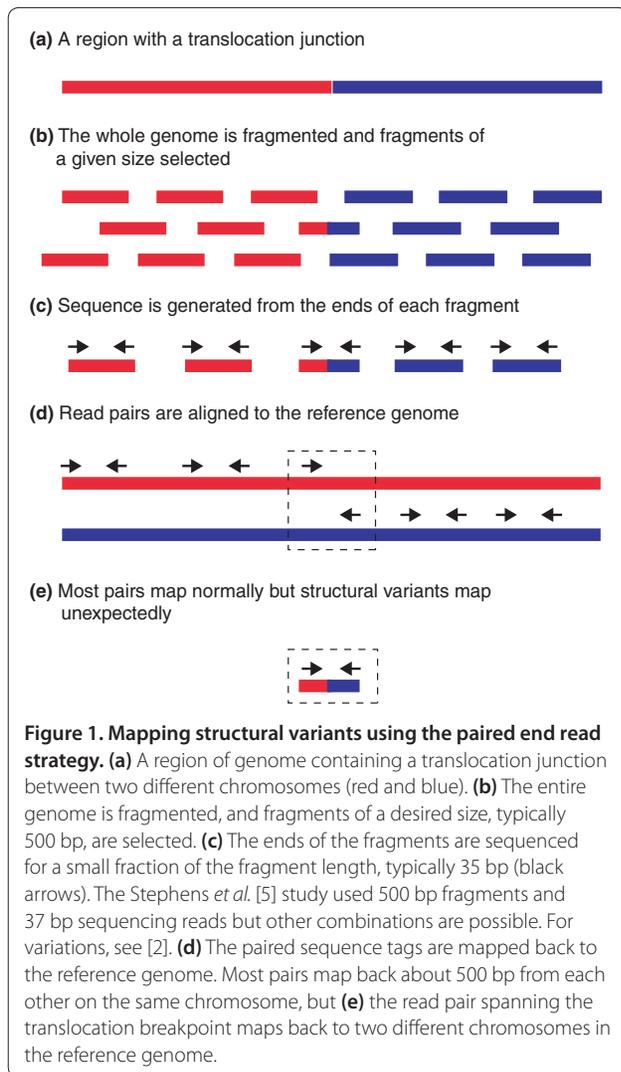
In stark contrast to leukemias, lymphomas and sarcomas, in which many important oncogenes have been identified at translocation breaks, we have a poor understanding of how structural variations contribute to carcinogenesis in common epithelial tumors [1,2]. Although we have relatively good knowledge of which genes can be point-mutated, amplified or deleted in these cancers, the sheer number and complexity of their genome rearrangements has made it difficult to identify genes at chromosome breakpoints [2]. We have known for several years that recurrent gene fusions are found in common epithelial cancers, following the discovery of the *TMPRSS2-ERG* and related fusions in prostate cancer [3] and *EML4-ALK* in lung cancer [4]. However, these fusions were discovered by essentially one-off methods and it remains to be seen whether these are isolated examples or the tip of an iceberg.

Stephens *et al.* [5] recently presented the first large-scale survey of somatically acquired structural variation in the genomes of cancers, with the explicit goal of discovering genes disrupted and fused at chromosome breakpoints. The authors [5] used massively parallel paired end sequencing to find genome rearrangements in 24 breast cancers - 9 of which were from immortal cell lines and 15 from primary tumors. Although these data pertain to breast cancer, we think many of the findings will also be relevant to other common cancers, and certainly they are consistent with a preceding pilot study of two lung cancer cell lines [6]. The Stephens *et al.* [5] study revealed that structural variants contribute significantly to the mutational burden of many breast cancers, but also that genes are often fused or otherwise disrupted by mechanisms we have, so far, not appreciated.

## Massively parallel paired end sequencing

Massively parallel sequencing techniques generate very large numbers of sequence reads, but the reads are generally much shorter than in traditional sequencing, typically only tens of base pairs. To use these short sequence 'tags' efficiently to find structural rearrangements, 'paired end read' strategies have been developed

\*Correspondence: sn353@cam.ac.uk; pawe1@cam.ac.uk  
Hutchison-MRC Research Centre and Department of Pathology, University of Cambridge, Hills Road, Cambridge, CB2 0XZ, UK



(also known as ‘mate pair’ and ‘end sequence profiling’ strategies; Figure 1) [6]. The genome is broken into DNA fragments of selected size, for example 500 base pairs (bp) [5], and a short sequence, for example 37 bp, is read from each end of each DNA fragment to give paired sequences. Most of the fragments are normal, and their paired reads map back to the reference genome about 500 bp apart and in the correct orientation. Structural variants are discovered when read-pairs map unexpectedly, for example to two different chromosomes (translocation), too far apart (deletion), or in the wrong orientation (tandem duplication or inversion) (Figure 1). Considerable bioinformatic processing is required to interpret the huge volume of sequence data, but millions of paired reads are pruned down to a hundred or so structural variants per tumor, most of which can be confirmed by PCR.

Stephens *et al.* [5] estimate that 50% of structural variations were detected in their study. This may seem

like a low figure but, as the authors showed, it was sufficient to identify hundreds of structural variants and tens of fusion genes. The main reason for missing structural variants was that the amount of sequencing was not enough to sample all rearrangements. Also, breakpoints flanked by repeats may have been missed because reads from repetitive regions are currently discarded. We expect the proportion of structural variants detected to increase in the future as more sequencing reads are generated, the reads used are longer, and bioinformatic analysis is refined.

### Rearrangements in breast cancers are more numerous than expected

There were many more structural variants than most in the field would have anticipated [5]. For cell lines, the median number of rearrangements per sample was 101 and ranged from 58 to 245. For the tumors, the median was 38 and ranged from 1 to 231. Approximately 85% were intrachromosomal and less than 2 Mb [5], which explains why earlier molecular cytogenetic approaches, such as spectral karyotyping, array comparative genomic hybridization (CGH) and array painting [7], underestimated the number of rearrangements. These aberrations would not have been visible in metaphase chromosomes and many were copy-number neutral or too small to have shown up in most array CGH experiments.

### Many fusion genes were predicted and several were expressed

Many of the structural changes that Stephens *et al.* [5] found juxtaposed the coding regions of two genes. An important observation, extending earlier studies [2,7,8], was that some breast cancers can express several fused genes. Stephens *et al.* [5] showed that 21 novel fusion genes were expressed and in frame so potentially produced a functional fusion protein. Allowing for the estimated 50% detection rate, this would equate to two functional fusion genes per case. Most of the fusion genes were of unknown function but several involved known or likely cancer genes, such as *ETV6*, which is a known target of translocations and encodes a member of the oncogenic Ets transcription factor family, and *EHF*, which also encodes an Ets family member. Some genes seemed to be rearranged in several of the 24 samples but no recurrent gene fusions were identified by fluorescence *in situ* hybridization (FISH) or RT-PCR in a larger second set of tumors [5]. This may simply be a reflection of the heterogeneity of breast cancer - the samples used were chosen to represent a range of different tumor subtypes - or it may be that aberrant expression of an important 3' gene can be driven by several different 5' fusion transcript partners, as happens, for example, to the Ets-related gene *ERG* in prostate cancers.

## Unanticipated classes of structural variation

An unexpected finding [5] was a number of somatically acquired tandem duplications, a kind of structural change that has rarely been detected until recently but is interesting because it can lead to gene fusion [9]. A tandem duplication occurs when a small region from 3 kb to greater than 1 Mb is duplicated, usually in a head-to-tail orientation. Some tumors showed a distinctly higher number of tandem duplications than the others, which led the authors [5] to suggest that they were generated by a specific repair defect. The *BRCA1* and *BRCA2* mutant tumors had fewer tandem duplications than average, so the aberrant mechanism was probably not related to these pathways.

The second surprising finding [5] was that many small tandem duplications, inversions and deletions were entirely within genes. In many cases this affected the exon structure at the transcript level and novel isoforms were observed. Some of these rearrangements were in putative oncogenes, such as the transcription-factor-encoding gene *RUNX1*, so it is plausible that oncogenic activation could have occurred by removing or reshuffling exons that encode a repressive protein domain. Well-characterized tumor suppressor genes such as the retinoblastoma gene *RB* also had internal rearrangements and it is possible these genes were inactivated through frame shift in the transcript or by removing important protein domains.

Two questions arise from these observations [5]: firstly, whether the roles of genes such as *RUNX1* and *RB* have been underestimated in breast cancer, because these kinds of mutation would not be detected by Sanger sequencing studies on individual coding exons; and secondly, whether there are numerous small rearrangements of this kind in other, karyotypically normal, cancers.

## Drivers and passengers?

It is remarkable how many mutations, whether sequence-level, epigenetic or structural, are now being discovered in cancer genomes [5,10,11]. Many are probably 'passenger' mutations, that is, random mutational noise, but some must be selected, 'driver' events and, as the number and variety of known mutations increases, estimates for the number of 'driving' mutations in cancer are tending to increase [2,12].

The problem of distinguishing driver and passenger mutations is as acute for structural mutations as it is for point mutations [10-13]. Stephens *et al.* [5] estimate that approximately 2% of genome rearrangements of the types they found would generate an in-frame fusion gene by chance. They observed 1.6%, which suggests that the majority of gene fusions, like the majority of point mutations, are not selected events.

## Conclusions

The Stephens *et al.* [5] study is the first indication that genome-wide structural analysis of a relatively large

number of samples, including primary tumors, is already an achievable goal. More importantly, it illustrates that such studies are worthwhile as they can create a large yield of new candidate oncogenes and tumor suppressor genes.

Clearly, the next step is to find genes or gene families that are recurrently fused or rearranged in a subset of tumors. Thanks to the methodologies and bioinformatic tools already validated by pilot studies [5,6] we can expect large surveys of several cancer types to appear within 2 or 3 years. This will allow us to address the question of recurrence and move on to establish the clinical relevance and potential for targeted intervention.

For the time being, massively parallel paired end sequencing will remain a research tool, but the basic cost of an analysis like that of Stephens *et al.* [5] is already down to a few thousands of euros per case, so it is conceivable that we will see it used in the clinic in the not too distant future. Indeed, while this article was in press, Velculescu and colleagues [14] announced a possible clinical application, using paired end reads to find a structural 'fingerprint' of a tumor that could be detected in the patient's serum and so used to monitor progression.

### Abbreviations

bp, base pair; CGH, comparative genomic hybridization.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

SN drafted the article; SN and PE edited and approved the manuscript.

### Acknowledgements

SN is supported by the UK Medical Research Council.

### Author information

SN is a graduate student and PE is university faculty in the Hutchison-MRC Research Centre and Department of Pathology, University of Cambridge, UK.

Published: 17 March 2010

### References

1. Mitelman F, Johansson B, Mertens F: **The impact of translocations and gene fusions on cancer causation.** *Nat Rev Cancer* 2007, **7**:233-245.
2. Edwards PAW: **Fusion genes and chromosome translocations in the common epithelial cancers.** *J Pathol* 2010, **220**:244-254.
3. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM: **Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer.** *Science* 2005, **310**:644-648.
4. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, Aburatani H, Niki T, Sohara Y, Sugiyama Y, Mano H: **Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer.** *Nature* 2007, **448**:561-566.
5. Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, Stebbings LA, Leroy C, Edkins S, Mudie LJ, Greenman CD, Jia M, Latimer C, Teague JW, Lau KW, Burton J, Quail MA, Swerdlow H, Churcher C, Natrajan R, Sieuwerts AM, Martens JW, Silver DP, Langerød A, Russnes HE, Foekens JA, Reis-Filho JS, van't Veer L, Richardson AL, Borresen-Dale AL, *et al.*: **Complex landscapes of somatic rearrangement in human breast cancer genomes.** *Nature* 2009, **462**:1005-1010.

6. Campbell PJ, Stephens PJ, Pleasance ED, O'Meara S, Li H, Santarius T, Stebbings LA, Leroy C, Edkins S, Hardy C, Teague JW, Menzies A, Goodhead I, Turner DJ, Clee CM, Quail MA, Cox A, Brown C, Durbin R, Hurlles ME, Edwards PA, Bignell GR, Stratton MR, Futreal PA: **Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing.** *Nat Genet* 2008, **40**:722-729.
7. Howarth KD, Blood KA, Ng BL, Beavis JC, Chua Y, Cooke SL, Raby S, Ichimura K, Collins VP, Carter NP, Edwards PA: **Array painting reveals a high frequency of balanced translocations in breast cancer cell lines that break in cancer-relevant genes.** *Oncogene* 2008, **27**:3345-3359.
8. Hampton OA, Den Hollander P, Miller CA, Delgado DA, Li J, Coarfa C, Harris RA, Richards S, Scherer SE, Muzny DM, Gibbs RA, Lee AV, Milosavljevic A: **A sequence-level map of chromosomal breakpoints in the MCF-7 breast cancer cell line yields insights into the evolution of a cancer genome.** *Genome Res* 2009, **19**:167-177.
9. Jones DT, Kocialkowski S, Liu L, Pearson DM, Bocklund LM, Ichimura K, Collins VP: **Tandem duplication producing a novel oncogenic BRAF fusion gene defines the majority of pilocytic astrocytomas.** *Cancer Res* 2008, **68**:8673-8677.
10. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, O'Meara S, Vastrik I, Schmidt EE, Avis T, Barthorpe S, Bhamra G, Buck G, Choudhury B, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jenkinson A, Jones D, *et al.*: **Patterns of somatic mutation in human cancer genomes.** *Nature* 2007, **446**:153-158.
11. Wood LD, Parsons DW, Jones S, Lin J, Sjöblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV, *et al.*: **The genomic landscapes of human breast and colorectal cancers.** *Science* 2007, **318**:1108-1113.
12. Stratton MR, Campbell PJ, Futreal PA: **The cancer genome.** *Nature* 2009, **458**:719-724.
13. Getz G, Hofling H, Mesirov JP, Golub TR, Meyerson M, Tibshirani R, Lander ES: **Comment on "The consensus coding sequences of human breast and colorectal cancers".** *Science* 2007, **317**:1500.
14. Leary RJ, Kinde I, Diehl F, Schmidt K, Clouser C, Duncan C, Antipova A, Lee C, McKernan K, De La Vega FM, Kinzler KW, Vogelstein B, Diaz LA Jr, and Velculescu VE: **Development of personalized tumor biomarkers using massively parallel sequencing.** *Sci Transl Med* 2010, **2**:20ra14.

doi:10.1186/gm140

**Cite this article as:** Newman S, Edwards PAW: **High-throughput analysis of chromosome translocations and other genome rearrangements in epithelial cancers.** *Genome Medicine* 2010, **2**:19.