RESEARCH HIGHLIGHT

Go retro and get a GRIP

Kim Wong, David J Adams* and Thomas M Keane

See related Research article: http://genomebiology.com/2013/14/3/R22

Abstract

Gene retrocopy insertions are a source of new genes and new gene functions, and can now be identified using paired-end whole genome sequencing data.

Large eukaryotic genomes are littered with the remains of past invaders. Gene retrocopies, which are generated from reverse transcription of mRNA from source genes, have previously been described in mammals and fruit flies [1,2]. Surprisingly, a number of these intronless gene retrocopies are functional, implicating gene retrotransposition as a possible mechanism of evolution. Two recently published studies have revealed gene retrocopies hiding within the genomes of humans, chimpanzees and mice [3,4]. In this issue of *Genome Biology*, Adam Ewing and colleagues describe how they analyzed the patterns of paired-end read mapping data from whole-genome sequences to identify gene retrocopies present in one or more individuals, but absent from the reference genome [4]. They refer to these as gene retrocopy insertion polymorphisms, or GRIPs.

Although GRIPs may have deleterious effects, they may also play an important role in adaptive evolution. Unlike a new gene that evolves from the duplication of a DNA segment, a GRIP is generally inserted at a locus distant from its source gene, where it may be placed under the control of new regulatory elements [5]. In turn, this may allow expression of the gene retrocopy in new cell types or in response to new stimuli. Thus, the inserted retrocopy may evolve a completely new gene function. Alternatively, it may be possible for a GRIP to help buffer against loss of a source gene, essentially storing a gene copy in reserve for when it may be required. It is also possible that a GRIP may amplify gene dosage. That said, as GRIPs appear to be distributed relatively randomly throughout vertebrate genomes, the expectation is that

many are located in repressive chromatin environments, and thus their expression is silenced [4].

Grappling with GRIP maps

How many GRIPs are there in vertebrate genomes? To address this question, Ewing and colleagues developed GRIPper, a software tool that detects gene retrocopy insertions by finding distinct discordant mapping patterns from paired-end sequence data. GRIPper identifies clusters of mapped reads for which corresponding mate pairs are mapped discordantly to the exons of a source gene; this strategy is similar to that used previously to find non-reference retroelement insertions in human and mouse genomes [6,7]. To add further confidence to their calls, the authors carried out a local *de novo* assembly around candidate breakpoints at GRIP insertions. Using whole-genome sequencing data from a subset of individuals from the 1000 Genomes Project, 10 chimpanzee genomes from the PanMap project and 17 mouse genomes from the Mouse Genomes Project, GRIPper was used to catalog a collection of 48 distinct GRIPs in humans, 19 GRIPs in chimpanzees and 755 GRIPs in mice.

The 48 gene retrocopies found in humans were identified in subsets of 1,024 individuals, indicating that these events are polymorphic. As expected for a heritable genomic polymorphism, a number of the GRIPs are restricted to individuals from defined geographical areas. The authors estimate a rate of 1 novel heritable GRIP per 5,177 individuals, which is comparable to the rate of 1 novel heritable GRIP in 6,804 individuals in the chimpanzee. In the mouse, where the 17 genomes analyzed by GRIPper span approximately 2 million years of evolution, between 100 and 200 GRIPs were found in the wild-derived inbred strains, which are most divergent from the reference, while on average only 56 were found in the genomes of the common laboratory strains, which are largely *Mus musculus domesticus*-derived. The sharing of GRIPs largely recapitulated the ancestral origins of these mice.

To determine whether gene retrocopy insertions also occur in the soma, Ewing and colleagues examined data from six cancer types from The Cancer Genome Atlas (TCGA). Intriguingly, by comparing tumor samples with

*Correspondence: da1@sanger.ac.uk
Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1HH, UK

their matched normal samples, evidence of somatic gene retrocopy insertions for three genes was revealed.

Functional consequences of GRIP

What are the functional characteristics of GRIP source genes? Predictably, the genes that are the source of pseudogenes, gene copies already present in the reference genome that have lost their protein-coding capacity or are otherwise no longer expressed in the cell, overlap with those for new GRIPs.

In particular, highly and widely expressed genes with ribosomal functions, genes involved in metabolic processes, signal transduction genes and transcriptional regulators are enriched as source genes for GRIPs. This may suggest that gene expression levels or possibly transcript stability may influence the likelihood that extra copies of a gene are inserted into the genome as a GRIP.

Importantly, several of the source genes of GRIPs found in the human genomes analyzed are known to have multiple copies elsewhere in the reference genome. For example, *POLR2C*, *HSPE1* and *SNRPC* (encoding an RNA polymerase II component, a heat shock protein and a component of the U1 small nuclear ribonucleoprotein particle, respectively) are GRIPs in persons of African descent, and *COX7C* (a cytochrome *c* oxidase gene), *NACC1* (a transcription factor gene) and three ribosomal genes, *RPL22*, *RPS2* and *RPL37A*, are GRIPs in persons of Chinese/Japanese descent. Intriguingly, RNA interference knockdown experiments have revealed potential roles for several of these genes, namely *POLR2C*, *HSPE1* and *SNRPC*, in cell viability, and for *NACC1* in the cell cycle [8]. *RPS2* and *RPL37A* knockdown experiments have also revealed important roles for these genes in fundamental biological processes. For example, knockdown of *RPL37A* has been linked to nucleolar pre-40S maturation defects, 60S biogenesis defects and decreased viability [8]. Thus, the source genes for GRIPs also appear to be enriched for essential genes.

In cancers, several genes were found to be retrotransposed insertions, including *MYH11*, encoding a major contractile protein, and *GAS5*, encoding a regulator of growth arrest. The high similarity of retrotransposed elements to their source genes may provide a mechanism for amplification of oncogene dosage. Indeed, *MYH11* fusions with the gene *CBFB* (core binding factor B) are found in acute myeloid leukemia (AML), where they are particularly associated with a distinct clinical subclass of the disease called M4Eo [9]. In addition to dysregulation of *CBFB*, which is known to be a key aspect of M4Eo, dysregulation of *MYH11* may also contribute to M4Eo leukemogenesis. One possible mechanism for this is the association between myosin heavy chain 11 (MYH11) and histone deacetylase complex subunit Sin3a (mSin3A), a subunit of the corepressor complex, and with histone deacetylases, an event that has been linked to the repression of runt-related transcription factor 1 (RUNX1)-mediated gene regulation [9]. Dysfunction of *MYH11* has also been linked to other types of cancer, including prostate cancer. Since *GAS5*, which is a non-coding RNA gene, is critical for the normal growth arrest of both leukemic and untransformed human T lymphocytes, it also has potential as a cancer driver gene [10]. Since *GAS5* is important for the inhibitory effects of the chemotherapeutic rapamycin, somatic alterations of this gene might have important implications for therapy. Thus, somatic retrocopy insertions may represent an important yet underappreciated contributor to cancer evolution.

Future directions

With the help of GRIPer and the catalog of variants described in this study, we can begin to examine how evolution of humans and other species has been affected by gene retrotransposition, and to what extent GRIPs contribute to adaptive evolution. Now that it is possible to catalog somatic retrocopy insertions in cancer samples, functional studies to determine the contribution of these mutations to tumorigenesis may also be performed. Looking ahead, it is clear that software such as GRIPer should be included in the toolkit used by analysts to catalog structural variation in eukaryotes.

References
1. Zhang Z, Harrison PM, Liu Y, Gerstein M: **Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome.** *Genome Res* 2003, **13**:2541-2558.
2. Schrider DR, Stevens K, Cardeno CM, Langley CH, Hahn MW: **Genome-wide analysis of retrogene polymorphisms in *Drosophila melanogaster*.** *Genome Res* 2011, **21**:2087-2095.
3. Schrider DR, Navarro FC, Galante PA, Parmigiani RB, Camargo AA, Hahn MW, de Souza SJ: **Gene copy-number polymorphism caused by retrotransposition in humans.** *PLoS Genet* 2013, **9**:e1003242.
4. Ewing A, Ballinger T, Earl D, Platform BIGSaAPa, Harris C, Ding L, Wilson R, Haussler D: **Retrotransposition of gene transcripts leads to structural variation in mammalian genomes.** *Genome Biol* 2013, **14**:R22.
5. Schacherer J, Tourrette Y, Souciet JL, Potier S, De Montigny J: **Recovery of a function involving gene duplication by retroposition in *Saccharomyces cerevisiae*.** *Genome Res* 2004, **14**:1291-1297.
6. Keane TM, Wong K, Adams DJ: **RetroSeq: transposable element discovery from next-generation sequencing data.** *Bioinformatics* 2013, **29**:389-390.
7. Stewart C, Kural D, Strömberg MP, Walker JA, Konkel MK, Stütz AM, Urban AE, Grubert F, Lam HY, Lee WP, Busby M, Indap AR, Garrison E, Huff C, Xing J, Snyder MP, Jorde LB, Batzer MA, Korbel JO, Marth GT; 1000 Genomes Project: **A comprehensive map of mobile element insertion polymorphisms in humans.** *PLoS Genet* 2011, **7**:e1002236.
8. Schmidt EE, Pelz O, Buhlmann S, Kerr G, Horn T, Boutros M: **GenomeRNAi: a database for cell-based and in vivo RNAi phenotypes, 2013 update.**

Nucleic Acids Res 2013, **41:**D1021-1026.
9. Schwind S, Edwards CG, Nicolet D, Mrózek K, Maharry K, Wu YZ, Paschka P, Eisfeld AK, Hoellerbauer P, Becker H, Metzeler KH, Curfman J, Kohlschmidt J, Prior TW, Kolitz JE, Blum W, Pettenati MJ, Dal Cin P, Carroll AJ, Caligiuri MA, Larson RA, Volinia S, Marcucci G, Bloomfield CD; Alliance for Clinical Trials in Oncology: **inv(16)/t(16;16) acute myeloid leukemia with non-type A CBFB-MYH11 fusions associate with distinct clinical and genetic features and lack KIT mutations.** *Blood* 2013, **121:**385-391.

10. Williams GT, Mourtada-Maarabouni M, Farzaneh F: **A critical role for non-coding RNA GAS5 in growth arrest and rapamycin inhibition in human T-lymphocytes.** *Biochem Soc Trans* 2011, **39:**482-486.