

RAGE and PRAY

DJ Smith^{1,2}

¹Department of Molecular and Medical Pharmacology; ²Crump Institute for Molecular Imaging, UCLA School of Medicine, Los Angeles, CA, USA

Random activation of gene expression, or RAGE, is an innovative new approach to gene discovery. The method involves the creation of normalized transcript libraries *in vivo*, which also allows the phenotypic consequences of gene overexpression to be assessed. Thus, RAGE seamlessly integrates the traditionally distinct processes of gene discovery and functional assignment, giving pharmacogenomics a powerful new tool.

When an electrical engineer wants to test the workings of a digital circuit, the engineer can probe the circuit with electrical pulses and monitor the consequences. An analogous approach would be ideal for investigating the logic underlying the genetic circuitry of the cell, but until now, there has been no simple method for accomplishing this. In a recent paper in *Nature Biotechnology*, Harrington *et al* describe RAGE, an ingenious method that allows comprehensive activation of gene expression in mammalian cells.¹ The cells are transfected with a series of vectors containing a

constitutive promoter linked to an exon with an accompanying splice donor site (Figure 1). When the vector is randomly integrated into a single site in the genome, the promoter transcribes the vector encoded exon, which can then splice to downstream exons of adjacent genes. Genes activated by RAGE can be specifically detected using primers that recognize the RAGE vector and the relevant downstream gene using reverse transcription-polymerase chain reaction (RT-PCR). The multiplicity of RAGE vectors are employed to account for the three possible open reading frames of the endogenous gene, and also to confer various desirable new properties upon the fusion transcript. For example, one set of RAGE vectors gives the fusion protein a secretion signal sequence, which allows facile detection of secreted proteins.

A major stumbling block to other gene discovery methods, for example expressed sequence tags (ESTs) and serial analysis of gene expression (SAGE), is the fact that genes are transcribed at

vastly different rates. This imposes the law of diminishing returns, making discovery of genes transcribed at low levels difficult and time consuming. RAGE helps remove this obstacle, since the genes are all activated by the same promoter, and should therefore be transcribed with equal efficiency. In essence, RAGE provides a pathway to normalized production of transcripts, hence vastly improving the efficiency of gene discovery.

To test the RAGE concept, a library of 5 million clones was created and split into 200 pools, with each clone representing an independent RAGE integration event. Using RT-PCR, 19 genes were tested which were known to be silent in the parental cell line. Each of the 19 genes was found to be activated in at least one cell pool, suggesting that RAGE libraries of modest size can comprehensively activate the majority of genes in the genome. In addition, the tested genes are normally present in cDNA libraries at frequencies that vary over several orders of magnitude. The similar frequency of activation of these genes in the RAGE library indicated that the library was successfully normalized.

RAGE was then employed to see if it could identify new genes. Vector tagged cDNAs were produced from pools of transfected cells and sequenced to give RAGE sequence tags, or RSTs. Over 75 000 RSTs were acquired, and these could be placed into around 20 000 unique gene clusters. More than 80% of the identified genes were isolated only once, and over 90% were isolated no more than twice. These data strongly support the idea that RAGE libraries are normalized, and also suggest that integration hotspots are not a problem with the technology. Surprisingly, of the 20 000 unique genes, 53% were completely novel. To further assess the fidelity of RAGE, a careful study was undertaken of the unique RSTs that map to chromosome 22. This chromosome was chosen for analysis because it has been the previous subject of intense annotation efforts. In the RAGE analysis, 41 new chromosome 22 genes were discovered from the

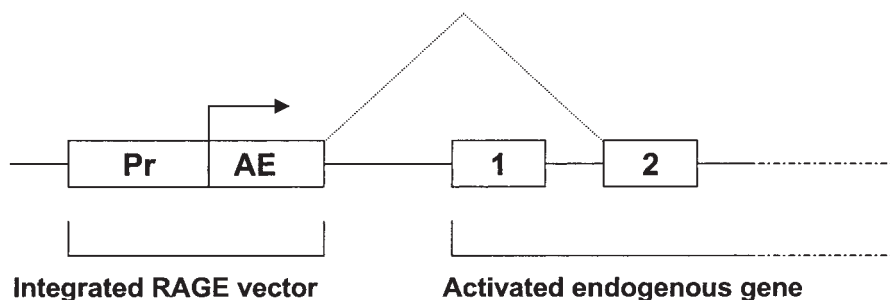


Figure 1 The principle of RAGE. The RAGE vector is randomly integrated into the genome via transfection. In the example shown in the figure, the RAGE vector has inserted into the genome upstream of an endogenous gene. The RAGE vector contains a constitutive promoter (Pr) and an activation exon (AE), the latter of which contains a start codon and a splice donor. When transcribed by the constitutive promoter, the activation exon splices to the second exon of the endogenous gene, making a fusion transcript.

RSTs. These novel genes had not previously been identified by gene prediction programs or EST methods. In addition, RSTs corresponding to known chromosome 22 genes showed correct splicing, implying good fidelity of RAGE. An additional experiment assessed over 800 novel genes predicted by RAGE throughout the genome. Using RT-PCR expression profiling of various tissues, more than 50% of the novel RSTs were verified as proper genes based on expression in at least one tissue. Interestingly, the verification rate of less than 100% using RT-PCR was most likely a reflection of the limitations of RT-PCR rather than RAGE, as a similar verification rate was found for known genes using RT-PCR.

Finally, one of the RAGE vectors had a dihydrofolate reductase (DHFR) cassette. This was employed to amplify identified RAGE insertion sites expressing human growth hormone (hGH) and human erythropoietin (hEpo) by using increasing concentrations of methotrexate. The strategy successfully produced cloned RAGE cell lines which synthesized large amounts of these commercially valuable proteins.

The major finding of RAGE is that despite intensive community-wide gene discovery efforts, an enormous number of genes have yet to be identified. This is a real surprise, and suggests that the ongoing debate over the total number of genes in the human genome will not be settled any time soon.² The use of RT-PCR expression profiles from various tissues provides very convincing evidence that most of the new genes identified by RAGE are real. However, it might be interesting to see what happens when a RAGE vector is inserted upstream of a known piece of 'junk' or non-coding DNA. In the future, a number of different variations on RAGE can be envisaged. For example, RAGE vectors with tetracycline inducible promoters could be constructed to better catalog genes that

are potentially toxic or that confer growth disadvantages when overexpressed.³

Like most new ideas, RAGE did not spring fully formed from nowhere, but rather has intellectual precursors. Conceptually similar attempts to discover phenotypes based on random gene overexpression have been performed in *Drosophila*.⁴⁻⁶ However, these screens were only moderate throughput, and were in a non-mammalian context. In addition, a number of other schemes have been employed for creation of normalized cDNA libraries, but these are technically demanding and often meet with limited success. RAGE is the first method in which random activation of genes has been efficiently accomplished at high throughput, allowing a new and more efficient approach to gene discovery using transcript normalization. Moreover, RAGE also permits functional genomics analyses. This can either be through phenotypes caused by overexpression or, when the RAGE vector inserts downstream of the first exon of an endogenous gene, due to the expression of a truncated fusion gene resulting in dominant negative phenotypes. Haploinsufficient phenotypes due to insertion of the RAGE vector are also a possibility. Thus, RAGE seamlessly combines two customarily separate domains of genomics, gene discovery and function, into one high throughput methodology.

What promise does RAGE hold for pharmacogenomics? First of all, of course, the discovery of the entire universe of genes helps pharmacogenomics by providing a full complement of genetic targets for study. But beyond this, RAGE should facilitate the linking of drugs with gene targets by, for example, observing altered drug effects in RAGE cell lines where a target gene product is overexpressed. Analogous strategies have been employed with great success in mutant yeast strains, and RAGE provides an efficient way

for achieving this goal in mammalian cells.⁷ The major challenge now facing RAGE, and indeed nearly all functional genomics approaches today, is how to efficiently and comprehensively acquire the phenotypic state of the genetically manipulated cell. I propose a new name for these novel, and as yet, unimagined phenotypic capture methods: phenotypic response assays, or PRAY. In this way, the field of genomics can look forward to a time where we will all be able to RAGE and PRAY.

DUALITY OF INTEREST

None declared.

Correspondence should be sent to:

DJ Smith, Department of Molecular and Medical Pharmacology, 23-120 CHS, Box 951735, UCLA School of Medicine, Los Angeles, CA 90095-1735, USA.

Tel: +1 310 206 0086

Fax: +1 310 825 6267

E-mail: DSmith@mednet.ucla.edu

REFERENCES

- 1 Harrington JJ, Sherf B, Rundlett S, Jackson PD, Perry R, Cain S *et al.* Creation of genome-wide protein expression libraries using random activation of gene expression. *Nat Biotechnol* 2001; **19**: 440-445.
- 2 Pennisi E. Human genome project—and the gene number is...? *Science* 2000; **288**: 1146-1147.
- 3 Mansuy IM, Bujard H. Tetracycline-regulated gene expression in the brain. *Curr Opin Neurobiol* 2000; **10**: 593-596.
- 4 Brand AH, Perrimon N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* 1993; **118**: 401-415.
- 5 Hay BA, Maile R, Rubin GM. P element insertion-dependent gene activation in the *Drosophila* eye. *Proc Natl Acad Sci USA* 1997; **94**: 5195-5200.
- 6 Rørth P, Szabo K, Bailey A, Lavery T, Rehm J, Rubin GM *et al.* Systematic gain-of-function genetics in *Drosophila*. *Development* 1998; **125**: 1049-1057.
- 7 Giaever G, Shoemaker DD, Jones TW, Liang H, Winzler EA, Astromoff A *et al.* Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nat Genet* 1999; **21**: 278-283.