Expression of human endogenous retroviruses

Merja Oja

Department of Computer Science, University of Helsinki
P.O.Box 68, 00014 University of Helsinki, Finland
merja.oja@cs.helsinki.fi

Abstract

The human endogenous retroviruses (HERVs) are remnants of ancient infections of retroviruses and they were previously thought to be inactive junk DNA. Now, HERV expression has been detected in both normal tissues and in diseased patients. This paper is a brief review of current laboratory and *in silico* methods used to analyze HERV expression, and the results achievable with these methods. The focus is in methods that analyze larger samples simultaneously, and have potential to produce data sets suitable for data mining.

1 Introduction

About eight per cent of human DNA consist of human endogenous retroviruses (HERVs) [4]. The HERVs are remnants of ancient infections of retroviruses[6]. In many recent studies it has been suggested that HERVs have an effect on the genome in which they reside, that they are not just junk DNA as earlier was believed. The reported functions include: the necessary expression of HERV genes in normal tissue, the possible HERV expression in diseased individuals and the regulation of human genes by HERV-derived promoters.

Some of the HERVs have retained their coding sequences, possibly due to purifying selection during evolution. This suggests that the conserved HERVs serve a beneficial role in our genome. Indeed, two HERV genes have been found to be expressed during an essential step in formation of the placenta [10, 9, 2]. In addition, HERVs are expressed in germ-line cells, and it has been hypothesized that they function in protection against their infective exogenous counterparts [7].

Recent studies have revealed that HERVs may play an active role in many diseases [13, 7]. Retroviral genes are expressed in schizophrenia [19], autoimmune diseases [16, 12] and cancer [11, 15, 1]. The exact role of the HERV genes in the formation of the disease is often unknown.

In order to understand the potential roles of HERVs in various diseases, we must first know how the HERVs are expressed in healthy tissues. For example, if HERV expression is observed in breast cancer, it is vital to know if similar expression could be detected in healthy breast tissue as well.

In this paper I will review the work done in HERV expression analysis and briefly describe three methods applied to measuring the expression. The focus is in methods that analyze larger samples simultaneously, and have potential to produce data sets suitable for data mining. Mainly this paper serves the purpose of a literature study preceding work on computational analysis of HERV expression.

2 Human endogenous retrovirus families

Human endogenous retroviruses are grouped into families. Sequences belonging to the same family are assumed to have originated from a single infection event. After the infection the retrovirus has proliferated by copying its sequence to multiple locations in the genome. The members of a family are very similar in their sequence but not identical. Mutations and rearrangements have happened between and after the copying events. Various sources list 22-45 different families [8, 5, 22].

3 Studying HERV expression

A primer/probe/query sequence matching one representative element of a HERV family will match also other sequences of the family. Various works have analyzed the expression of HERV families in this fashion.

3.1 RT-PCR

Reverse transcriptase polymerase chain reaction (RT-PCR) is a sensitive technique for mRNA detection and quantization. First, mRNA is isolated from the samples (tissues, cancer cell lines etc.). Then the mRNA is reverse transcribed from *HERV-family specific primers* in the RT-PCR reaction. The reaction amplifies the HERV sequences, and they can be sequenced later.

A HERV expression analysis with RT-PCR can produce quantitative information about the abundance of the expression in each analyzed sample. Furthermore, the expressed HERVs can be sequenced, and then mapped back to the genome or analyzed in some other fashion. For example, the sequencing can reveal whether one or several of the members in the analyzed HERV family are expressed.

The following works employ the RT-PCR method for the detection of HERV-related sequences. Yi *et al.* have analyzed the expression of HERV-W [24], HERV-S [26] and HERV-F [25] families in 12 healthy and 18 cancerous tissues, the PCR-products were also sequenced and analyzed. Expression of HERV-E family was analyzed in healthy and cancerous prostate cells in [23]. HERV-K expression was studied in healthy and cancerous cells, and the expressed sequences were mapped to the human genome in [21]. Retroviral complete open reading frames (ORFs) were searched from the human genome in [3], the expression of these ORFs was verified with RT-PCR.

These works show that different HERV families have different expression profiles. For example HERV-F expression is detected only in placenta (among tested healthy tissues), whereas HERV-W can be seen in almost any healthy tissue. Differences between healthy and cancerous tissues were also observed, giving support to the hypothesis that HERVs might cause disease.

Sequencing of RT-PCR products revealed that several elements in a family are expressed [25], and that these elements can have different expression profiles [24].

3.2 Retrovirus-specific microarray

Seifarth *et al.* have developed an oligonucleotide-based retrovirus-specific microarray. The microarray contains 54 retroviral probes, together these target 20 retroviral families. The construction and usage of the array is described in detail in [18]. The procedure of using the microarray is two phased. First, the mRNA of the sample is subjected to broadly targeting RT-PCR, which amplifies the retroviruslike sequences in the sample. Then the amplified (now cDNA) sequences are labeled with fluorescence and hybridized onto the microarray, similarly as in gene expression microarrays.

Currently the retrovirus-specific microarray is able to produce only yes or no type of information. It's advantage is that it measures this information for all families simultaneously, when for example the RT-PCR technique needs one reaction per family. Hopefully, the retrovirus-specific microarray technology improves in accuracy and specificity in the future, and allows the quantitative expression analysis of more specific groups of HERVs (subfamilies, individual integrations, etc.).

The retrovirus-specific microarray was used in [17] to analyze HERV expression in 19 healthy human tissues. Different families were found to have distinct expression patterns. Furthermore, there was variation among the tissues, most active members of HERV families were detected in skin, thyroid gland, placenta, and tissues of reproductive organs. The qualitative results obtained with the retrovirus microarray were confirmed with RT-PCR for a subset of families and tissues.

3.3 EST-based analysis

Expressed sequence tags (ESTs) are sequence fragments corresponding to mRNAs in analyzed cells. An EST-library is constructed in the following way. First, the mRNA is isolated from the sample and the mRNA sequences are reverse transcribed to cDNA. All the cDNAs are sequenced which results in the ESTs. These ESTs are then submitted to **GenBank** and eventually to **dbEST**, a database of ESTs.

The generation procedure of an ESTs is not error free, reverse transcription and/or sequencing steps can introduce errors and the ESTs can be truncated. One gene can produce several mRNA sequences, which leads to several ESTs originating from the same gene. **UniGene** is a database where the ESTs have been grouped to clusters, each representing one potential gene.

The EST databases contain a wealth of information and great potential for analyzing HERV expression *in silico*. The EST database can, for example, be searched for HERVs expressed in interesting tissues or for ESTs originating from specific HERV families. The ESTs also contain quantitative information on the amount of expression.

Stauffer *et al.* screened the **dbEST** with representative sequences for HERV-K(HML2), HERV-E, HERV-H and HERV-W families, resulting in 428 matches

[20]. They observed differential expression of different families, and also differences for different sequences in the HERV-K(HML2) family. The ESTs matching HERV-sequences originated both from normal and cancerous tissues.

Kim *et al.* used extensive database screening. First, they screened the whole human genome with RepeatMasker[1] to find retroviral sequences, resulting in ∼40,000 HERVs. Then they matched these HERVs to **dbEST** and **UniGene**, in order to identify expressed HERVs. Then they mapped the HERV-related **UniGene** clusters to gene ontologies in order to understand the possible cellular functions of HERVs.

Okahara *et al.* have performed EST database screens with representative HERV sequences to detect candidate expressed HERV loci in the human genome. These loci were verified with RT-PCR analysis [14]. In other words, they were doing (HERV) gene finding.

4 Conclusions and future work

The research reviewed above has shown that there is variation in the expression profiles of different HERV families. Furthermore, differential expression of members in HERV-K and HERV-W families were observed in RT-PCR studies. These results show that HERV sequences are expressed tissue and condition specifically and that the control is sequence specific.

To better understand the potential in expressed HERVs, it will be necessary to move from family level expression measurements to sequence level measurements. The upstream regions of the individual expressed sequences can then be studied and the control mechanisms understood.

Large scale study of sequence level expression is possible with EST analysis. The EST data is, however, very noisy and thus it is hard to determine which individual sequence in a family produced the EST. In the near future I will develop statistical models that are able to tackle this problem. The models will hopefully be able to estimate robustly the probability for the expression of an individual family member in each condition. The expression of the best candidates can then be verified with sensitive laboratory techniques.

References

[1] V. Armbruester, M. Sauter, E. Krautkraemer, E. Meese, A. Kleiman, B. Best, K. Roemer, and N. Mueller-Lantzsch. A novel gene from the human endogenous retrovirus K expressed in transformed cells. *Clinical Cancer Research*, 8(6):1800–7, 2002.

[2] S. Blaise, N. de Parseval, L. Bénit, and T. Heidmann. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proceedings of the National Academy of Sciences*, 100(22):13013–8, 2003.

[3] N. de Parseval, V. Lazar, c. C. Jean-Fran L. Benit, and T. Heidmann. Survey of human genes of retroviral origin: Identification and transcriptome of the genes with coding capacity for complete envelope proteins. *Journal of Virology*, 77(19):10414–22, 2003.

[4] E. Lander *et al.* Initial sequencing and analysis of the human genome. *Nature*, 409:860–921, 2001.

[5] R. Gifford and M. Tristem. The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes*, 26(3):291–315, 2003.

[1] RepeatMasker is a program that screens DNA sequences for interspersed repeats and low complexity DNA sequences. It can be found from http://www.repeatmasker.org/

[6] D. J. Griffiths. Endogenous retroviruses in the human genome sequence. *Genome Biology*, 2(6):reviews1017.1–1017.5, 2001.

[7] R. Löwer. The pathogenic potential of endogenous retroviruses: facts and fantasies. *Trends in Microbiology*, 7(9):350–56, 1999.

[8] D. L. Mager and P. Medstrand. *Encyclopedia of the Human Genome*, chapter Retroviral repeat sequences. Nature Publishing Group, 2004.

[9] S. Mi, X. Lee, X. ping Li, G. M. Veldman, H. Finnerty, L. Racie, E. LaVallie, X.-Y. Tang, P. Edouard, S. Howes, J. C. Keith Jr., and J. M. McCoy. Suncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature*, 403:785–9, 2000.

[10] A. Muir, A. Lever, and A. Moffett. Expression and functions of human endogenous retroviruses in the placenta: An update. *Placenta*, 25(Suppl. 1):S16–S25, 2004.

[11] T. Muster, A. Waltenberger, A. Grassauer, S. Hirschl, P. Caucig, I. Romirer, D. Födinger, H. Seppele, O. Schanab, C. Magin-Lachmann, R. Löwer, B. Jansen, H. Pehamberger, and K. Wolff. An endogenous retrovirus derived from human melanoma cells. *Cancer Research*, 63:8735–41, 2003.

[12] K. Nakagawa and L. C. Harrison. The potential roles of endogenous retroviruses in autoimmunity. *Immunological Reviews*, 152:193–236, 1996.

[13] P. N. Nelson, P. R. Carnegie, J. Martin, H. Davari Ejtehadi, P. Hooley, D. Roden, S. Rowland-Jones, P. Warren, J. Astley, and P. G. Murray. Demystified ... human endogenous retroviruses. *Molecular Pathology*, 56:11–18, 2003.

[14] G. Okahara, S. Matsubara, T. Oda, J. Sugimoto, Y. Jinno, and F. Kanaya. Expression analyses of human endogenous retroviruses (HERVs): tissue-specific and developmental stage-dependent expression of HERVs. *Genomics*, 84:982–90, 2004.

[15] S. Patzke, M. Lindeskog, E. Munthe, and H. C. Aasheim. Characterization of a novel human endogenous retrovirus, HERV-H/F, expressed in human leukemia cell lines. *Virology*, 303(1):164–73, 2002.

[16] J. L. Portis. Perspectives on the role of endogenous human retroviruses in autoimmune diseases. *Virology*, 296:1–5, 2002.

[17] W. Seifarth, O. Frank, U. Zeifelder, B. Spiess, A. D. Greenwood, R. Hehlmann, and C. Leib-Mösch. Comprehensive analysis of human endogenous retrovirus transcriptional activity in human tissues with a retrovirus-specific microarray. *Journal of Virology*, 79(1):341–52, 2005.

[18] W. Seifarth, B. Spiess, U. Zeifelder, C. Speth, R. Hehlmann, and C. Leib-Mösch. Asessment of retroviral activity using a universal retrovirus chip. *Journal of Virological Methods*, 112:89–91, 2003.

[19] B. S. Shastry. Schizophrenia: a genetic perspective (review). *International Journal of Molecular Medicine*, 9(3):207–12, 2002.

[20] Y. Stauffer, G. Theiler, P. Sperisen, Y. Lebedev, and C. V. Jongeneel. Digital expression profiles of human endogenous retroviral families in normal and cancerous tissues. *Cancer Immunity*, 4(2), 2004.

[21] J. Sugimoto, N. Matsuura, Y. Kinjo, N. Takasu, T. Oda, and Y. Jinno. Transcriptionallly active HERV-K genes: Identification, isolation and chromosomal mapping. *Genomics*, 72:137–44, 2001.

[22] M. Tristem. Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. *Journal of Virology*, 74(8):3715–30, 2000.

[23] F. Wang-Johanning, A. R. Frost, B. Jian, R. Azerou, D. W. Lu, D.-T. Chen, and G. L. Johanning. Detecting the expression of human endogenous retrovirus E envelope transcripts in human prostate adenocarcinoma. *Cancer*, 98(1):187–97, 2003.

[24] J.-M. Yi, H.-M. Kim, and H.-S. Kim. Expression of the human endogneous retrovirus HERV-W family in various human tissues and cancer cells. *Journal of General Virology*, 85:1203–10, 2004.

[25] J.-M. Yi and H.-S. Kim. Expression analysis of endogenous retroviral elements belonging to the HERV-F family from human tissues and cancer cells. *Cancer Letters*.

[26] J.-M. Yi, T.-H. Kim, J.-W. Huh, K. S. Park, S. B. Jang, H.-M. Kim, and H.-S. Kim. Human endogenous retroviral elements belonging to the HERV-S family from human tissues, cancer cells, and primates: expression, strucure, phylogeny and evolution. *Gene*, 342:283–92, 2004.