

doi: 10.3969/j.issn.1672-5565.2014.02.07

基于生物信息学的 HERV 研究现状与发展趋势

韩九强, 吕红强*, 刘俊, 张善新

(西安交通大学电子与信息工程学院, 西安 710049)

摘要:近年来研究证实,人内源性逆转录病毒(HERV)中末端重复序列(LTR)和囊膜蛋白基因(*env*)与基因异常表达关系密切。本文阐述了 HERV 国内外研究现状与发展动态。从生物信息学角度对 LTR 中启动子、多聚腺苷化(PolyA)信号和增强子等顺式作用元件以及 *env* 中的免疫抑制区(ISD)进行了重点分析。指出上述元件和潜在功能区在基因的异常表达和囊膜蛋白的免疫抑制中起关键作用,必将成为 HERV 与基因异常表达关系研究的热点与趋势。通过此类研究,以期对癌症疾病的早期诊断和治疗靶位点的选择奠定理论基础。

关键词:人内源性逆转录病毒 长末端重复序列 启动子 多聚腺苷化信号 免疫抑制区

中图分类号:TP183 **文献标志码:**A **文章编号:**1672-5565(2014)-02-117-06

Bioinformatics research and development of HERV

HAN Jiuqiang, LÜ Hongqiang*, LIU Jun, ZHANG Shanxin

(School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Recent studies have confirmed that long terminal repeats (LTR) and the envelope gene (*env*) of human endogenous retroviruses (HERV) are closely related to aberrant gene expression. In this paper, the situations and development trends of domestic and foreign research on HERV were elaborated. The promoter, polyadenylation (PolyA) signal, enhancer and other cis-acting elements of LTR as well as immunosuppressive domain (ISD) of *env* in HERV were analyzed in the perspective of bioinformatics. It was suggested that the elements and potential functional domain, which play important roles in the abnormal gene expression and immunosuppression of envelope proteins, will become the focuses and trends of future research on the relationship between HERV and abnormal gene expression. It is hoped that these studies will be helpful to theoretical foundation for early diagnosis and selection of treatment target site of cancer.

Keywords: Human endogenous retroviruses; Long terminal repeats; Promoter; Polyadenylation signal; Immunosuppressive domain

世界卫生组织 WHO 在 2013 年发布的调查报告中指出,2008 年全球死于癌症的人数多达 760 万, 约占所有死亡人数的 13%, 预计到 2030 年全球将有超过 1 310 万人死于癌症^[1]。美国癌症学会估计的数据表明,2012 年美国共有超过 163 万例新发癌症病例,其中有约 58 万死亡病例^[2]。据保守估计,2013 年欧盟国家的癌症死亡人数预计将超过 130 万^[3]。《2012 中国肿瘤登记年报》指出,我国的肿瘤死亡率高达 180.54/10 万,每年癌症死亡人数

高达 270 万。近 20 年来癌症呈现年轻化,且发病率和死亡率呈走高趋势。预计到 2020 年,我国癌症新发病例将达到 300 万^[4]。因此,对于人类癌症预防、肿瘤早期诊断、治疗方法、药物开发、肿瘤分子标记物和癌症组织特异靶位点识别预测等问题的研究给生物医学界提出了新的重大挑战。

尽管大量医学研究已经证实人内源性逆转录病毒 HERV (Human Endogenous Retrovirus) 与癌症疾病的发生发展有重要关系^[5],但实验分析手段注释

收稿日期:2013-12-10;修回日期:2014-01-10.

资助项目:国家自然科学基金(61105021)资助.

作者简介:韩九强,男,教授,博士生导师,研究方向:机器视觉,生物信息学;E-mail:jqhan@mail.xjtu.edu.cn.

*通信作者:吕红强,男,硕士研究生,研究方向:生物信息学;E-mail:lhqxinghun@stu.xjtu.edu.cn.

HERV 中全部的重要元件与位点面临困难,进一步研究人内源性逆转录病毒 HERV 与癌症发生发展的相互关系几乎不可能。然而,生物信息学为此提供了新的分析手段和方法。目前,人内源性逆转录病毒 HERV 序列中的元件与位点注释及其与癌症的相互关系已成为生物信息学研究的热点问题^[6-7]。

1 HERV 国内外研究现状及发展动态

1.1 国内外研究现状

人内源性逆转录病毒 HERV 是在几百万年前整合到人类基因组中的一种逆转录病毒在经过遗传变异后产生的一种残余物,约占整个人类基因组的 8%^[8]。HERV 大多以非典型的单 LTR 形式存在,而典型 HERV 的完整序列结构如图 1 所示。其两端分别有一个长末端重复序列 (Long terminal

repeats, LTR), 中间是结构基因和逆转录因子结合区域 (PBS, PPT)^[9]。LTR 中含有启动子、多聚腺苷化 (Polyadenylation signal, PolyA) 信号和增强子等顺式作用元件。结构基因包括 *gag*, *pro*, *pol* 和 *env* (envelope gene)。长期的进化过程使 HERV 失去了病毒蛋白的编码功能。但近年来大量医学实验研究发现,HERV 导致的基因异常表达与人类诸多疾病,特别是癌症的发生和发展存在关系^[9-10]。目前,已在结肠癌、乳腺癌、黑瘤素瘤、生殖细胞肿瘤等严重影响人类健康的癌症组织中都发现了 HERV 家族元件的异常表达^[5, 11-12]。虽然目前对 HERV 的准确致癌机制尚不清楚,但研究普遍认为, LTR 的顺式作用元件对周围基因有异常调控作用。囊膜蛋白基因 *env* 的编码结果对细胞融合与癌症转移产生作用,它的免疫抑制功能是导致 HERV 致癌的重要因素^[6-7]。

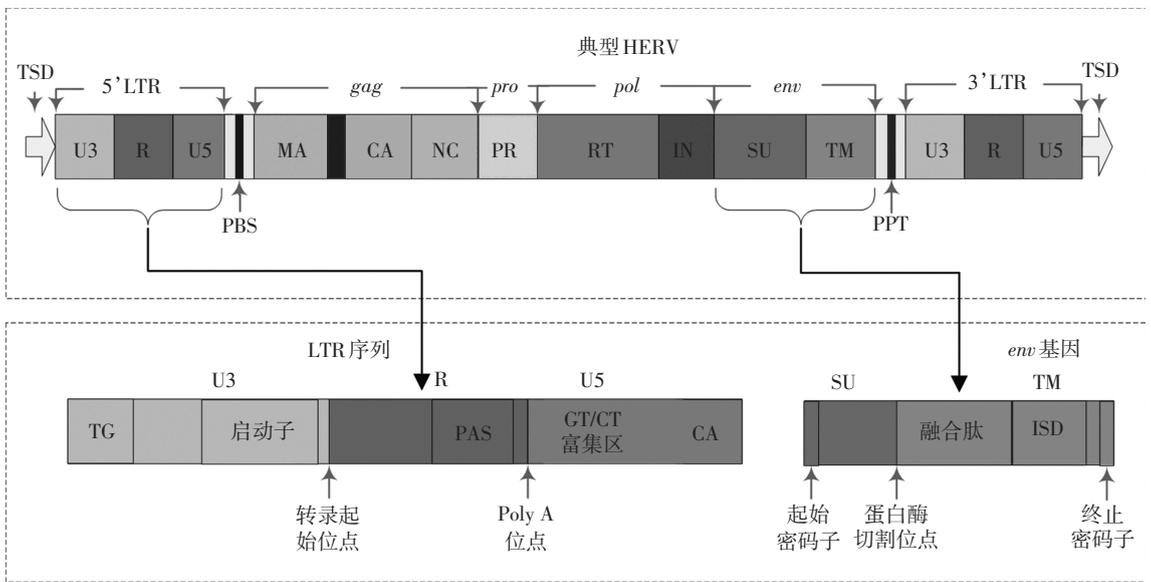


图 1 典型 HERV 序列结构框图

Fig.1 Block diagram of the typical HERV sequence

HERV 是外源性逆转录病毒感染人类生殖细胞后,整合到人类基因组中,并经过长期的遗传变异形成的人类基因组中广泛存在的一种逆转录转座子^[13],它具有分布范围广、数量多、种类繁多、保守性差、以及结构复杂等特点。HERV 几乎在人类基因组的每一条染色体上都有分布,人类基因组中有超过 10 万条的拷贝^[14]。HERV 约有 3 个不同属至少 31 个家族^[15]。由于突变导致其保守性差,形成了不完整的病毒基因组的复杂结构。不同家族 HERV 的组成结构变化多样,内部存在多种顺式作用元件。目前,HERV 的研究主要集中在 HERV-K、HERV-W、HERV-E、HERV-H 和 HERV-FRD 等少

数家族中,因为在正常或者患者的蛋白质组中已经发现了此类家族的蛋白质产物。对各家族之间基因异常表达调控关系的研究还未见报道。研究 HERV 及其基因异常表达是一个极具挑战性的科学难题,也是一个基础性的科学问题。基于实验分析的传统方法研究 HERV 及其基因异常表达问题费时费力,而生物信息学为此提供了新的分析手段和方法,这也是后基因组时代人类基因研究注释的发展趋势之一。

采用生物信息学方法对 HERV 研究主要集中在对完整结构 HERV 识别方面。2003 年,美国佐治亚大学 McCarthy 和 McDonald 通过识别一定距离范围内相同的 LTR 的方法识别 LTR 逆转录转座

子^[16];2007年,瑞典乌普萨拉大学 Sperber 等人提出在对 LTR 逆转录转座子保守序列打分基础上,通过重构内部结构对其进行识别^[17];2009年,德国汉堡大学 Steinbiss 等人提出采用 *denovo* 方法对 LTR 进行定位研究,采用识别内部基因的 pHMM 方法识别 LTR 逆转录转座子^[18];2012年, Dashti 等人采用支持向量机 SVM 方法识别 LTR 逆转录转座子^[19];2009年,瑞典乌普萨拉大学 Benachenhou 等人提出采用 HMM 方法对不同属内源性逆转录病毒 LTR 进行识别研究^[20]。近年来,许多新的生物信息学方法被用来注释基因及其内部功能区域。2011年, Morey 等人采用基于 DNA 自由能方法对拟南芥和水稻基因的启动子进行比较性识别研究^[21];2013年, Zhou 等人对伪密码子进行离散小波变换,将其变换结果作为 SVM 的输入参数识别基因的启动子^[22];2009年, Ahmed 等人采用核苷酸频率法结合 SVM 方法进行基因 PolyA 位点识别研究^[23];2012年, Kalkatawi 等人采用多特征和随机森林相结合方法进行 PolyA 位点识别研究^[24];2012年, Zhang 和 Yan 采用经验模式分解和傅里叶变换相结合方法进行外显子预测研究^[25];2012年, Zhang 等人采用改进的统计最优空滤波器(SONF)方法进行蛋白编码区域预测研究^[26]。

在生物信息学领域,许多学者基于基因表达谱数据,对基因与癌症疾病的关系进行了研究。2011年, Zheng 等人通过采用稀疏表征方法对微阵列基因数据进行处理,对基因组中与癌症相关的基因进行研究^[27];2012年, Perot 等人制作了具有40个正常人和39个肿瘤患者的5573个人内源性逆转录病毒 HERV 的微阵列芯片,通过统计学方法分析了 HERV 的 LTR 启动子和 PolyA 信号与癌症的关系^[28]。研究表明,内源性逆转录病毒 LTR 中的启动子和多聚腺苷化信号和囊膜蛋白基因 *env* 与癌症疾病的发生发展有紧密关系。

国内在人内源性逆转录病毒 HERV 方面的研究较晚,但发展迅速。目前,浙江大学,武汉大学,天津医科大学等单位在 HERV 研究方面深入并取得重要进展。其中浙江大学郑树教授团队研究发现了与直肠癌有关的3个 HERV 新基因;武汉大学朱帆教授对精神分裂症与 HERV 的关系研究有新的建树;天津医科大学顾林主任医师团队研究了 HERV 对乳腺癌发生发展的影响。此外,清华大学李衍达院士团队、复旦大学郝柏林院士团队、天津大学张春霆院士团队、厦门大学吉国力教授、内蒙古大学李前忠教授等在基因、剪接位点、基因表达谱分析、多聚腺苷化位点识别等方面均展开了深入研究。

1.2 国内外研究发展动态

从2001年人类基因组测序完成以来,在全基因组范围内研究人内源性逆转录病毒 HERV 相关内容的时间跨度约10年。10年来,关于人内源性逆转录病毒 HERV 相关概念从提出到目前的研究发展动态分析如下:

(1)2001年“人类基因组分析”论文在《Nature》发表后,基因组序列研究逐步展开,随着人类基因组测序完成,研究进入后基因组时代;

(2)2003年提出识别 LTR 逆转录转座子,2012年仍有相关研究论文;

(3)2004年发现 LTR 的启动子具有组织特异性,但至今未见有造成 LTR 启动子特异性原因研究方面的报道;

(4)2007年提出采用模式识别方法识别 *env*,但至今仍未见采用模式识别方法对其内部区域研究的报道。

(5)2007年提出研究 HERV 与癌症关系,2012年该方面内容研究增多;

(6)2008年开始研究 HERV 家族与癌症关系,2012年相关研究内容增多;

(7)2009年提出采用生物信息学方法研究 LTR 的 PolyA 位点,至今未见针对 HERV 中 LTR 的 PolyA 位点识别方法的研究报道;

(8)2012年继续有 LTR 逆转录转座子相关问题的研究;

(9)2013年有关人内源性逆转录病毒与癌症关系方面研究,已有报道。

上述分析表明,人内源性逆转录病毒及其与癌症关系的研究已成为学术界关注的热点问题。HERV 国内外研究发展动态如图2所示。

2 HERV 研究热点与趋势分析

2.1 LTR 顺式作用元件

人内源性逆转录病毒 HERV 的长末端重复序列 LTR 中存在与基因转录相关的启动子、PolyA 信号和增强子等多种顺式作用元件,大量实验已经证实了这些顺式作用元件能影响其附近基因的表达^[9]。Mark 等证实了 HERV-P 启动子在人类睾丸细胞中调控12%神经元凋亡抑制蛋白的转录,而在肾脏组织中并未发现^[29];Philippe 等在 HERV 中鉴别出了326条具有启动子活性和209条具有 PolyA 活性的 LTR^[30];Ruda 等人在人睾丸癌细胞 Tera-1 中发现了人类特有的单 LTR(L47334)的表达增强,该表达增强源于 LTR 的增强子活性^[31];Rita 等人给

出了广义 LTR 转座子与人类基因组中顺式作用元件的调控信息统计数据^[32](见图 3)。很明显,启动子、可变剪接、PolyA 信号和增强子在 LTR 和人基因

组中具有明显的基因调控作用。因此,HERV 中 LTR 顺式作用元件的进一步分析与识别将会成为 HERV 研究的热点。

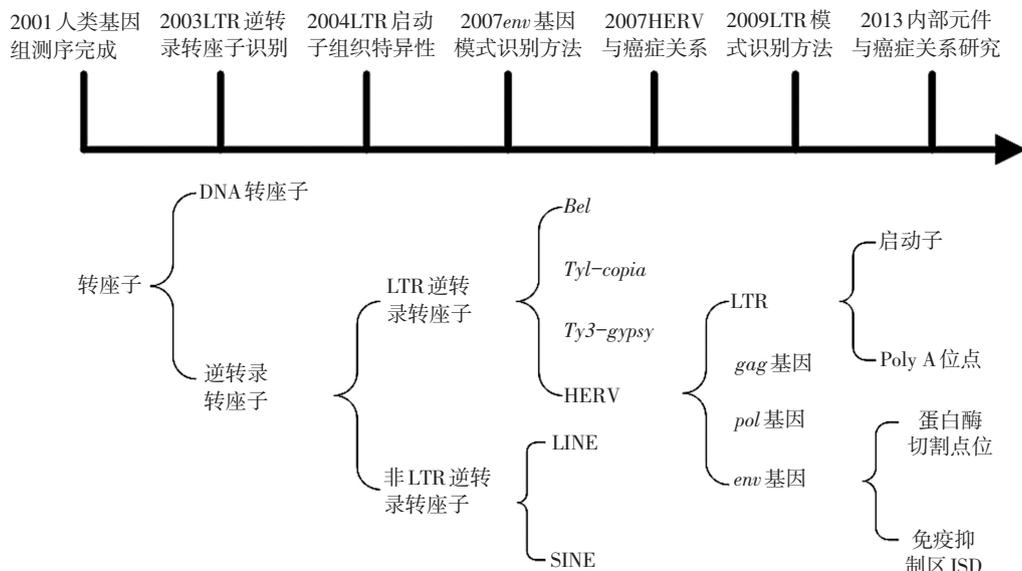


图 2 HERV 研究发展动态框图

Fig.2 Chart of development of HERV research

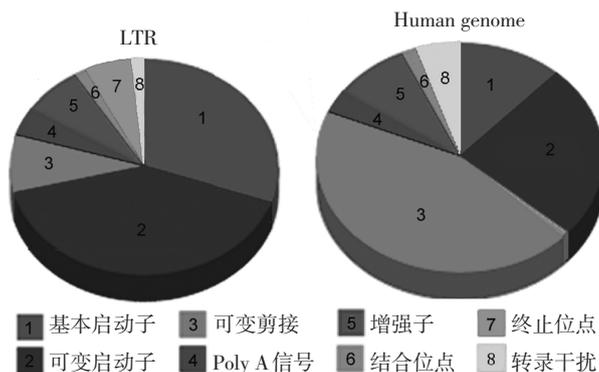


图 3 LTR 转座子与人染色体中顺式作用元件调控作用

Fig.3 Proportional regulation of cis-acting elements of LTR transposons and the human genome

2.2 env 免疫抑制区

囊膜蛋白基因 *env* 具有膜融合作用,可编码包膜蛋白,它是病毒入侵宿主细胞的关键。大量实验研究表明,*env* 的异常表达与多种癌症疾病相关。Wang-Johanning 等人发现,乳腺癌细胞表面 HERV-K 的 *env* 编码的蛋白表达明显高于良性肿瘤细胞^[33];Hahn 等人在黑色素瘤患者血液发现了 HERV-K 的 *env* 编码的蛋白产物^[34];Serafino 等人在研究中也发现 HERV-K 的 *env* 蛋白参与了黑色素瘤细胞的恶性转移^[35];Flockerzi 等人在生殖细胞肿瘤中检测到了 HERV-K(HML-2)的 *env* 蛋白抗体存在^[36];Feng 等人在卵巢癌患者的组织中检测到了 HERV3、HERV-E 和 HERV-K 家族 *env* 转录的

mRNA^[37]。研究发现,囊膜蛋白基因 *env* 的一个重要致癌机制是其编码的包膜蛋白和跨膜蛋白能够与宿主细胞融合,进而引起宿主细胞的严重免疫抑制,从而诱发产生肿瘤。而引起免疫抑制的关键蛋白来自跨膜蛋白中的免疫抑制区 (ISD)^[38-39]。因此,针对 HERV 结构基因中关键蛋白的研究,也将会加深对 HERV 编码区结构功能的进一步认识。

2.3 HERV 与异常表达关系

研究 HERV 的序列结构及其形成和发展过程,从基因调控网络水平理解 HERV 在基因异常表达过程中的角色和作用。筛选癌症组织中特异表达的 HERV,运用现有公用基因数据库大量的 HERV 序列数据、基因表达谱数据以及相应的注释数据,基于

现代生物信息学方法,分析揭示 HERV 与基因异常表达关系。

3 结论与展望

近年来研究证实,HERV 中 LTR 和 *env* 与基因异常表达关系密切。LTR 中的启动子、多聚腺苷化信号和增强子等顺式作用元件以及 *env* 中的免疫抑制区在基因的异常表达和囊膜蛋白的免疫抑制中起关键作用。因此,采用生物信息学方法,对上述元件和潜在功能区进行分析识别必将会成为 HERV 研究的热点问题。通过此类研究,揭示 HERV 中 LTR 和 *env* 与基因异常表达之间的关系,从而为癌症疾病的早期诊断和治疗靶位点的选择奠定理论基础。

参考文献(References)

- [1] WHO. Cancer [EB/OL]. <http://www.who.int/mediacentre/factsheets/fs297/en/index.html>, 2013-03-15/2013-12-01.
- [2] SIEGEL R, NAISHADHAM D, JEMAL A. Cancer Statistics, 2012 [J]. CA: a Cancer Journal for Clinicians, 2012, 62(1): 10-29.
- [3] MALVEZZI M, BERTUCCIO P, LEVI F, et al. European cancer mortality predictions for the year 2013 [J]. Annals of Oncology: Official Journal of the European Society for Medical Oncology / ESMO, 2013, 24(3): 792-800.
- [4] 郝捷, 陈万青. 中国肿瘤登记年报 [M]. 北京: 军事医学科学出版社, 2012.
HAO Jie, CHEN Wanqing. Chinese cancer registry annual report [M]. Beijing: Military Medical Science Press, 2012.
- [5] LEE E, ISKOW R, YANG L, et al. Landscape of somatic retrotransposition in human cancers [J]. Science, 2012, 337(6097): 967-971.
- [6] MULLINSC S, LINNEBACHER M. Human endogenous retroviruses and cancer: Causality and therapeutic possibilities [J]. World Journal of Gastroenterology, 2012, 18(42): 6027-6035.
- [7] CEGOLON L, SALATA C, WEIDERPASS E, et al. Human endogenous retroviruses and cancer prevention: evidence and prospects [J]. BMC Cancer, 2013, 13: 4-4.
- [8] LANDERE S, LINTON L M, BIRREN B, et al. Initial sequencing and analysis of the human genome [J]. Nature, 2001, 409(6822): 860-921.
- [9] JERNP, COFFIN J M. Effects of retroviruses on host genome function [J]. Annual Review of Genetics, 2008, 42: 709-732.
- [10] SINGH S K. Endogenous retroviruses: suspects in the disease world [J]. Future Microbiology, 2007, 2(3): 269-275.
- [11] RUPRECHTK, MAYER J, SAUTER M, et al. Endogenous retroviruses and cancer [J]. Cellular and Molecular Life Sciences, 2008, 65(21): 3366-3382.
- [12] KOWALCZYK M J, DANCZAK-PAZDROWSKA A, SZRAMKA-PAWLAK B, et al. Expression of selected human endogenous retroviral sequences in skin and peripheral blood mononuclear cells in morphea [J]. Archives of Medical Science, 2012, 8(5): 819-825.
- [13] LEVINH L, MORAN J V. Dynamic interactions between transposable elements and their hosts [J]. Nature Reviews Genetics, 2011, 12: 615-627.
- [14] BELSHAW R, KATZOURAKIS A, PACES J, et al. High copy number in human endogenous retrovirus families is associated with copying mechanisms in addition to reinfection [J]. Molecular Biology and Evolution, 2005, 22(4): 814-817.
- [15] JERNP, SPERBER G O, BLOMBERG J. Use of endogenous retroviral sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy [J]. Retrovirology, 2005, 2: 50.
- [16] MCCARTHY E M, McDonald J F. LTR_STRUC: A novel search and identification program for LTR retrotransposons [J]. Bioinformatics, 2003, 19(3): 362-367.
- [17] SPERBER G O, AIROLA T, JERN P, et al. Automated recognition of retroviral sequences in genomic data-RetroTector [J]. Nucleic Acids Research, 2007, 35(15): 4964-4976.
- [18] STEINBISS, WILLHOEFT U, GREMME G, et al. Fine-grained annotation and classification of de novo predicted LTR retrotransposons [J]. Nucleic Acids Research, 2009, 37(21): 7002-7013.
- [19] DASHTI T H, MASOUDI N A, ZARE F. Finding exact and solo LTR-Retrotransposons in biological sequences using SVM [J]. Iranian Journal of Chemistry & Chemical Engineering-International English Edition, 2012, 31: 111-116.
- [20] BENACHENHOU F, JERN P, OJA M, et al. Evolutionary conservation of orthoretroviral long terminal repeats (LTRs) and ab initio detection of single LTRs in genomic data [J]. PloS one, 2009, 4(4): e5179.
- [21] MOREY C, MOOKHERJEE S, RAJASEKARAN G, et al. DNA free energy-based promoter prediction and comparative analysis of arabidopsis and rice genomes [J]. Plant Physiology, 2011, 156(3): 1300-1315.
- [22] ZHOU Xuan, LI Zhanchao, DAI Zong, et al. Predicting promoters by pseudo-trinucleotide compositions based on discrete wavelets transform [J]. Journal of Theoretical

- Biology, 2013, 319: 1–7.
- [23] AHMED F, KUMAR M, RAGHAVA G P. Prediction of polyAdenylation signals in human DNA sequences using nucleotide frequencies [J]. *In Silico Biology*, 2009, 9(3): 135–148.
- [24] KALKATAWI M, RANGKUTI F, SCHRAMM M, et al. Dragon polyA spotter: predictor of poly (A) motifs within human genomic DNA sequences [J]. *Bioinformatics*, 2012, 28(1): 127–129.
- [25] ZHANG Weifeng, YAN Hong. Exon prediction using empirical mode decomposition and fourier transform of structural profiles of DNA sequences [J]. *Pattern Recognition*, 2012, 45: 947–955.
- [26] ZHANG Lei, TIAN Fengchun, WANG Shiyuan. A modified statistically optimal null filter method for recognizing protein-coding regions [J]. *Genomics Proteomics & Bioinformatics*, 2012, 10(3): 166–173.
- [27] ZHENG Chunhou, ZHANG Lei, NG T Y, et al. Metasample-based sparse representation for tumor classification [J]. *IEEE-ACM Transactions on Computational Biology and Bioinformatics*, 2011, 8(5): 1273–1282.
- [28] PEROTP, MUGNIER N, MONTGIRAUD C, et al. Microarray-based sketches of the HERV transcriptome landscape [J]. *PLoS One*, 2012, 7(6): e40194.
- [29] ROMANISH M T, LOCK W M, VAN D E, et al. Repeated recruitment of ltr retrotransposons as promoters by the anti-apoptotic locus naip during mammalian evolution [J]. *PLoS Genet*, 2007, 3(1): e10.
- [30] PEROT P, MUGNIER N, MONTGIRAUD C, et al. Microarray-based sketches of the herv transcriptome landscape [J]. *PLoS One*, 2012, 7(6): e40194.
- [31] RUDA V M, AKOPOV S B, TRUBETSKOY D O, et al. Tissue specificity of enhancer and promoter activities of a HERV-K(HML-2) LTR [J]. *Virus Research*, 2004, 104(1): 11–16.
- [32] REBOLLO R, FARIVAR S, MAGER D L. C-GATE-catalogue of genes affected by transposable elements [J]. *Mobile DNA*, 2012, 3(1): 1–9.
- [33] WANG-JOHANNING F, RYCAJ K, PLUMMER J B, et al. Immunotherapeutic potential of anti-human endogenous retrovirus-K envelope protein antibodies in targeting breast tumors [J]. *Journal of the National Cancer Institute*, 2012, 104(3): 189–210.
- [34] HAHN S, UGUREL S, HANSCHMANN K M, et al. Serological response to human endogenous retrovirus K in melanoma patients correlates with survival probability [J]. *Aids Research and Human Retroviruses*, 2008, 24(5): 717–723.
- [35] SERAFINO A, BALESTRIERI E, PIERIMARCHI P, et al. The activation of human endogenous retrovirus K (HERV-K) is implicated in melanoma cell malignant transformation [J]. *Experimental Cell Research*, 2009, 315(5): 849–862.
- [36] FLOCKERZI A, RUGGIERI A, FRANK O, et al. Expression patterns of transcribed human endogenous retrovirus HERV-K (HML-2) loci in human tissues and the need for a HERV transcriptome project [J]. *BMC Genomics*, 2008, 9: 354.
- [37] WANG-JOHANNING F, LIU J, RYCAJ K, et al. Expression of multiple human endogenous retrovirus surface envelope proteins in ovarian cancer [J]. *International Journal of Cancer*, 2007, 120(1): 81–90.
- [38] TOLOSA J M, SCHJENKEN J E, CLIFTON V L, et al. The endogenous retroviral envelope protein syncytin-1 inhibits LPS/PHA-stimulated cytokine responses in human blood and is sorted into placental exosomes [J]. *Placenta*, 2012, 33(11): 933–94.
- [39] LÜ Hongqiang, HAN Jiuqiang, LIU Jun, et al. ISDTool: A computational model for predicting immunosuppressive domain of HERVs [J]. *Computational Biology and Chemistry*, 2014, 49: 45–50.