UDC 618.146-006.52-036.2(09) https://doi.org/10.15407/biotech14.05.074

DETERMINING PROBABILITY OF CANCER CELL TRANSFOMATION AT HUMAN PAPILLOMAVIRUS INFECTION

L. P. Buchatskyi
V. V. Stcherbyc

Taras Shevchenko Kyiv National University, Ukraine

E-mail: iridolpb@gmail.com

Received 28.07.2021
Revised 18.10.2021
Accepted 30.10.2021

Aim. The purpose of the work was to assess the probability of cancerous transformation of cells for viruses of high and low oncogenic risk.

Method. It was used the statistical ensembles to determine the probability of cancer cell transformation.

For analysis there were selected oncoproteins E6 that directly influence the process of cancer cell transformation.

Results. Using normalized squared error (NSE) for viruses of high (20 strains) and low (153 strains) oncogenic risk, rank statistic of 2-exponential type was build. For productive papillomavirus infection, NSE function was determined as the growing accurate 2-exponent of a cell layer basal to the epithelial surface. Logarithm of NSE numerical values is proportional to the cell entropy that is connected with the availability of virus DNA. To calculate entropy, generalized Hartley formula was used with the informational cell of dimension d: $H = N^d LOG(NSE)$, where N is the generalized cell coordinate.

Conclusions. Using a statistical ensemble of E6 proteins separately for viruses of high and low oncogenic risk made it possible to assess the probability of cancerous transformation of cells, which was proportional to the ratio of the area of entropy of cancer transformation to the area of the productive entropy region papillomavirus infection.

Key words: human papillomavirus infection, carcinogenesis, cumulative Hartley entropy.

Human papillomaviruses are a large group of DNA viruses that, besides different forms of infections, may lead to the cancer degeneration of epithelial cells. At present, it is shown [1–4] that the presence of human papillomavirus is the necessary condition for development of cervical carcinoma. However, the papillomavirus infection alone is insufficient for neoplastic cell transformation; additional factors should take part [5–12]: high sexual activity, HIV infection, smoking, alcohol, virus loading, etc. Penetration of the virus takes place in the basal layer of epithelial cells and is accompanied with the cell proliferation. The assembly of papillomaviruses may occur only in mature cells of surface epithelial layers. Malignant neoformation occurs after rather long persistence of viruses in the basal layer of epithelial cells — the virus may be in the inactive, latent state for years. This is because the basal cells are under constant impact of the mature multilayer epithelium and incapable of cell division without the influence of external factors.

The human papillomavirus has an icosahedral protein coat 55 nm in diameter, which contains a minichromosome with double-helical DNA having ~ 8 000 bp [13, 14]. The virus replication takes place in the cell nucleus. As differentiation of epithelial cells is going on, DNA replication and expression of virus early proteins (E1, E2) take place. Late proteins (L1, L2) are produced at the final stage of virus assembly. So complicated scheme of the virus particle maturation ensures almost complete invulnerability of the virus for the immune system.

In the human organism, the papillomavirus may exist in the episome form (circular DNA) [15], which is typical for the productive infection [16], when oncoproteins E6 and E7 are not produced, but intact virus particles

are produced that may be integrated into the cell chromosome (integrated form) [17, 18]. Virus integrated form is capable of inducing malignant transformation wherein oncoproteins E6 and E7 are synthesized; interaction of the last with cell regulatory proteins p53 and pRb leads to deregulation of the cell cycle, which results in cancer cell regeneration [19, 20].

Virus infection course has several forms. The latent course has been determined as the virus persistence in basal epithelial layer [21, 22]. The virus is in the episome form and cannot lead to any pathology. Productive infection is accompanied by clinical infectious presentations (papillomas, warts, condylomas) and intensified cell growth in basal epithelial layer. At that, virus DNA is replicating in the infected cells.

Neoplasia occurs when DNA virus integrates into cell genome. At that, changes may happen in the epithelial cell structure of surface layers. Cell nucleus takes irregular shape; vacuoles appear in the cytoplasm. Most often, the injuries are localized in the transitional zone of cervix uteri at the interface of multilayer pavement and cylindrical epithelium.

In most cases of invasive tumor carcinoma, the virus exists in cells in the integrated form. Typical for malignancy atypical cells appear.

By their transforming activity towards the epithelial cells, all human papillomaviruses may be divided into two groups: papillomaviruses of high carcinogenic risk (HPV 16, 18, 26, 31, 33, 35, 39, 45, 51–53, 56, 58, 59, 66, 68, 70, 73, 82, 85); remainder (more than 150 strains) are papillomaviruses of low carcinogenic risk (HPV 6, 11, 40, 42–44, 54, 61, 62, 71, 72, 81, 83, 84 et al.) [23–25].

Papillomaviruses of low carcinogenic risk show themselves as the productive form of infection. Papillomaviruses of high carcinogenic risk (especially HPV16, 18) may lead to cancer of cervix uteri. There have been observed cases of virus infection spontaneous elimination, which causes are unknown [26, 27].

Separate types of HPV have less than 90% nucleotide identity in gene sequences of E6, E7, L1 [28]. Structural features of E6 and E7 oncoproteins are not so marked to be used for determining their oncogenic potential. Infection with viruses of high carcinogenic risk does not lead by itself to cancer of cervix uteri and disappears in most cases within 1–2 years. However, long virus persistence is essential (as main risk factor) for development of cancer cell transformation [29, 30]. Currently, development of cervix uteri cancer is described as the multistage process [31].

Accurate description the infectious process leading to cancer of cervix uteri is a rather complex problem. Therefore, different mathematical models of papillomavirus infection as a biological phenomenon have been proposed. Collaboration of mathematicians and biologists gives the possibility to evaluate conceptions of carcinogenesis and develop the generalized approach to treatment of the infectious processes.

Several deterministic models, based on differential equations [32–38], were proposed to describe different aspects of papillomavirus infection, including the potential effect of vaccination against HPV infection [39–42]. These models had only numerical solutions and could make it possible to analyze parameters leading to global equilibrium in spread of the infection.

To estimate the impact of additional factors on the infection, models with large numbers of intermediate states were suggested. Markov chains are at the basis of these models — both continuous and discrete [43–47]. In semi-Markov models, the probability to transfer into other state depends on the holding time in the initial state that is more realistic.

Papillomaviruses demonstrate strict species-specificity. Nevertheless, the investigation of papillomavirus infection on animals, especially on mice [48–51], may give more accurate information about the infectious process, which is necessary for developing new treatment methods and effective vaccines against HPV.

HPV of both high and low oncogenous risk have the same life cycle, but they differ greatly in the extent of operating cell cycle. This fact should be taken into account when using generalized approach to the structure of virus oncoproteins, since oncogenicity of the viruses cannot be determined by their real structures without implication of experimental data.

Proposed here method of statistical ensembles may significantly strengthen structural differences between oncoproteins of viruses with high and low oncogenous risk.

Material and Methods

To model processes connected with entropy changes of cells infected with papillomavirus, a statistical ensemble of E6 proteins was build separately for viruses of high and low oncogenic risk. The quantity of different amino acid residues in E6 proteins was

ranked according to their decreasing and approximated with exponential function. For each E6 protein, normalized squared error (NSE) of approximation, which is one of the main characteristic of the protein, was determined. Information on proteins of human papillomaviruses taken from the NCBI database (http://www.ncbi.nlm.nih.gov/genome/).

Results and Discussion

Statistical ensembles of oncoproteins E6

For analysis there were selected oncoproteins E6 that directly influence the process of cancer cell transformation. Protein E7 has short amino acid sequence insufficient for the structure analysis. Structural differences between proteins E6 in different papillomaviruses are not much obvious; therefore, statistical approach should be applies to the whole oncoproteins of two different papillomaviruses groups — with high and low oncogenic risk. Statistical approach gives the summarized characteristic of papillomavirus infection: function of virus spreading, function of spontaneous regression of viral infection, possibility of calculating the probability of cancer cell transformation.

Behavior of proteins E6 for different papillomaviruses may be generalized on basis of the statistical ensemble ("population") formed with using structural features of the proteins. Thereto, quantitative composition of amino acids may be arranged in alphabetical single-letter system of amino acid conventional signs. Obtained at that density of amino acid distribution has a rather complex polynomial approximation, typical for each oncoprotein E6. Such presentation reflects incidental differences of amino acid sequences in proteins E6, but it has no clear physical interpretation.

Let suppose that quantitative content of amino acids in the protein E6 for different viruses is arranged in decreasing order with reiterations. This will make it possible to select a maximally simple ordering function, which will exactly be the function of statistical ensemble. Analysis shows that three functions are relevant for oncoproteins E6: simple exponent, 2-exponent and linear function. To construct the ensemble, only one of them should be selected. For example, if simple exponent is chosen as the main function of ensemble, then the set of exponents will determine Gibbs statistical ensemble.

For each oncoprotein E6, the amino acid content was ranged in decreasing order. Then, using the method of least-squares, three functions were determined that best approximated the dependence of amino acid quantitative composition Q(Am) on the number of ordering (rank) N:

$$\text{simple exponent } Q(Am) = Q_{max}\,EXP\{-\alpha(N-1)\};$$

$$\text{2-exponent } Q(Am) = Q_1\,EXP\{-\alpha_1(N-1)\} + (Q_{max} - Q_1)\,EXP\{-\alpha_2(N-1)\};$$

$$\text{linear function } Q(Am) = Q_{max} - \alpha(N-1).$$

Hyperbolic function $Q(Am) = p/(N + m)^{\beta}$, proposed in [52], gives Q(Am) values close to the ones determined with the use of 2-exponent.

Papillomaviruses having E6 protein with linear Q(Am) characteristic: 7, 21, 25, 34, 36, 55, 67, 72, 73, 76, 80, 179.

Papillomaviruses having E6 protein with exponential Q(Am) characteristic: 16, 6, 35, 73, 43, 97, 40, 49, 5, 75, 76, 43, 41, 89, 102, 15, 20, 29, 110, 118, 132, 135, 139, 143, 144, 156, 172, 174.

E6 protein in the rest of papillomaviruses has 2-exponential characteristic.

To build the ensemble, for all proteins we have chosen the simple exponential dependence Q(Am) on the number of ordering, in spite of the fact that for some proteins such function gives no the best result with the use of least-squares approximation.

The Fig. 1 gives characteristic dependences Q (Am) on the number of ordering.

In the case of simple exponent, for each E6 oncoprotein, normalized squared error (NSE) of Q(Am) approximation is:

$$NSE\,(E6) = (1/e_p)(1/Q_{max})\,\Sigma_N\,[Q(N) - Q_{max}\,EXP\{-\alpha(N-1)\}]^2,$$

where $e_p = 10^{-4}$ is quantization step; Q(N) is the quantity of amino acids depending on the number of ordering.

For viruses of both high oncogenic risk (20 viruses) and low oncogenic risk (153 viruses), NSE function has the same dimension as Q(Am). Therefore, NSE is a generalized function (of some concentration). The function NSE may be considered as "the width" of α-level. It is significant that NSE values are unique for each E6 oncoprotein.

Generalized function NSE of E6 oncoproteins

Separately, for viruses of high and low oncogenic risk there was built NSE dependence of E6 oncoproteins on the number of reversed ranking. NSE function is a generalized function of papillomavirus DNA, which contains E6 gene.

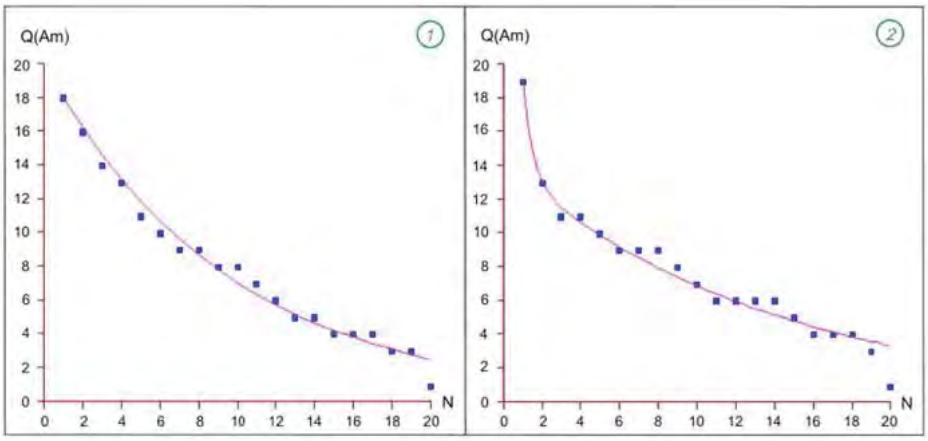

Fig. 1. 1 — **Amino acid content in E6 oncoprotein of HPV16 virus and its approximation with exponent:**
$Q(Am) = 18\ EXP\{-0.1043248(N-1)\}$
2 **—Amino acid content in E6 oncoprotein of HPV82 virus and its approximation with 2-exponent:**
$Q(Am) = 13.2\ EXP\{-0.073(N-1)\} + 5.8\ EXP\{-2.049177(N-1)\}$

NSE function may be considered as decreasing with time concentration of papillomavirus DNA at spontaneous elimination of virus infection. As increasing function, NSE represents the growth of virus loading at approaching to the epithelium border.

Decreasing NSE function is always a simple exponent, so as elimination of virus DNA is an irreversible process, irrespective of whether cancer cell transformation takes place or no, and therefore it has a maximally simple functional form. It should be pointed out that disappearance of virus DNA and its incorporation into the cell genome is described by an exponentially damped curve as expected at spontaneous elimination of virus DNA.

As epithelial cells become mature and move towards the epithelial surface, production of virus particles intensifies. That is why quantitative growth of virus DNA, and, accordingly, E6 genes, from the basal to surface layer follows exponential law (space form of the exponent). Spontaneous disappearance of the virus infection follows time exponential law.

Figs. 2 and 3 shows NSE plots, separately for viruses of high and low oncogenic risk. NSE function for viruses of high oncogenic risk differs little from simple exponent:

$$NSE_{HRSE} = 35.265\ EXP\{-0.1047024(N-1)\}.$$

As a generalized function, NSE is proportional to the probability of E6 oncoprotein decomposition.

Exact arrangement of E6 virus proteins at 2-exponential dependence of NSE requires experimental data. Generalized function determines only the image of NSE.

If 2-exponente is chosen as a main ensemble function, NSE function for viruses of high oncogenic risk will be reduced by half, and for viruses of low oncogenic risk — three fold decreased. Arrangement of E6 oncoproteins with NSE function will also change.

Integrated plot of NSE function for all analyzed papillomaviruses is shown on the Fig. 4. One may see from the plot that structural characteristics of E6 oncoproteins for both high and low oncogenous risk are mutually confused.

Hartley cumulative entropy

At productive infection, the quantity of virus DNA continuously increases when getting nearer to the epithelial surface. This process leads to the entropy growth in cells, since total information in them grows. Inverse process is spontaneous elimination of viruses. At cancer, cell entropy decreases, as there is no virus production. Intermediate forms of the infection may take place, but the main event leading to cancer is the incorporation of virus DNA into cell genome.

For productive virus infection, NSE function is defined as exact 2-exponent increasing from a basal cell layer to the epithelial surface. Decreasing cancer NSE function is the rough exponent, as events in a cell cannot be reversed. Logarithm of numerical values of NSE function is proportional to the cell entropy associated with the presence of virus DNA.

An integer value of NSE may be defined as the number of parameters (symbols) that characterize properties of E6 protein. Then LOG(NSE) is Hartley entropy (information measure) [53].

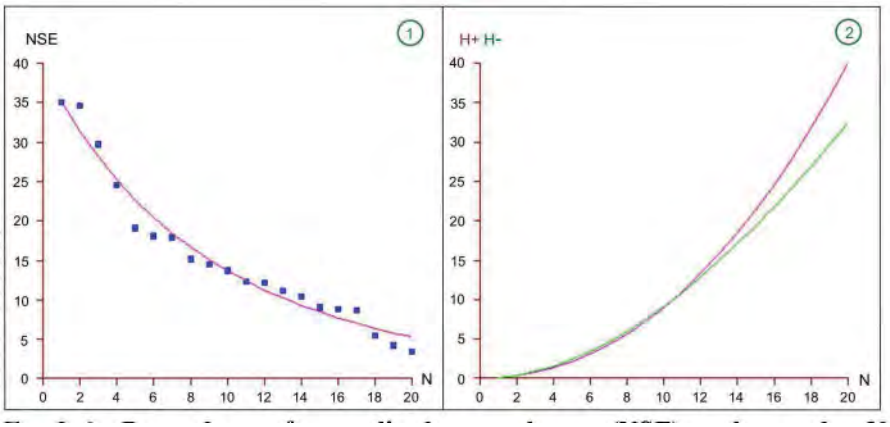

***Fig. 2. 1* — Dependence of normalized squared error (NSE) on the number N of reverse ranking of viruses with high oncogenic risk and exponential ap-proximation:**

$NSE_{HR} = 28.4\ EXP\{-\ 0.088999(N - 1)\} + 6.865\ EXP\{-\ 0.2197541(N - 1)\}$.

Arrangement of papillomaviruses with high oncogenic risk in decreasing order:

18, 45, 82, 59, 70, 53, 68, 56, 33, 73, 66, 85, 58, 51, 52, 31, 39, 26, 35, 16.

***2* — Plots of H^+ and H^- functions of cumulative entropy $d = 2$ for viruses of high oncogenic risk, N = 1...20:**

$H^+(N) = N^d LOG\ [28.4\ EXP\{-\ 0.088999(20 - N)\} + 6.865\ EXP\{-\ 0.2197541(20 - N)]$.

$H^-(N) = N^d LOG\ [35.265\ EXP\{-\ 0.1047024(N - 1)\}]$. Crosspoint of H^+ and H^- is at $N_0 = 10$.

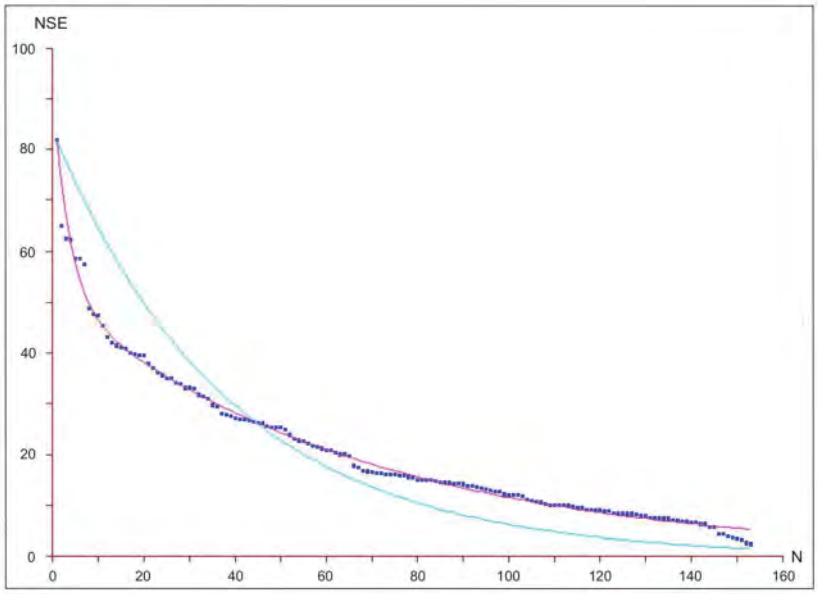

Fig. 3. **Dependence of normalized squared error (NSE) on the number N of reverse ranking of low oncogenous risk viruses:**

exponential approximation (lilac-coloured curve)

$NSE_{LR} = 50.5\ EXP\{-\ 0.0149(N - 1)\} + 31.352\ EXP\{-\ 0.2812865(N - 1)\}$

rough simple exponent (blue curve) is given by

$NSE_{LRSE} = 81.852\ EXP\{-\ 0.02603004(N - 1)\}$

arrangement of papillomaviruses of low oncogenic risk in decreasing order:

96, 204, 131, 129, 47, 197, 126, 171, 202, 140, 122, 121, 112, 105, 169, 1, 159, 115, 134, 141, 95, 151, 125, 22, 153, 113, 99, 9, 150, 63, 30, 111, 136, 201, 98, 104, 86, 199, 149, 96, 204, 131, 129, 47, 197, 126, 171, 202, 140, 122, 121, 112, 105, 169, 1, 159, 115, 134, 141, 95, 151, 125, 22, 153, 113, 99, 9, 150, 63, 30, 111, 136, 201, 98, 104, 86, 199, 149, 154, 23, 209, 109, 142, 48, 163, 117, 37, 180, 123, 3, 28, 145, 42, 17, 4, 87, 165, 20, 166, 173, 128, 179, 78, 168, 90, 60, 152, 69, 119, 100, 67, 175, 158, 12, 24, 184, 107, 65, 147, 50, 200, 32, 91, 11, 205, 170, 157, 161, 124, 138, 83, 21, 27, 167, 172, 81, 137, 133, 71, 49, 156, 120, 2, 74, 178, 92, 106, 44, 162, 94, 80, 38, 43, 36, 10, 34, 19, 84, 143, 77, 97, 61, 174, 76, 54, 110, 132, 7, 15, 144, 139, 130, 135, 164, 41, 5, 25, 88, 6, 75, 62, 55, 57, 72, 118, 89, 40, 114, 29, 93, 13, 102.

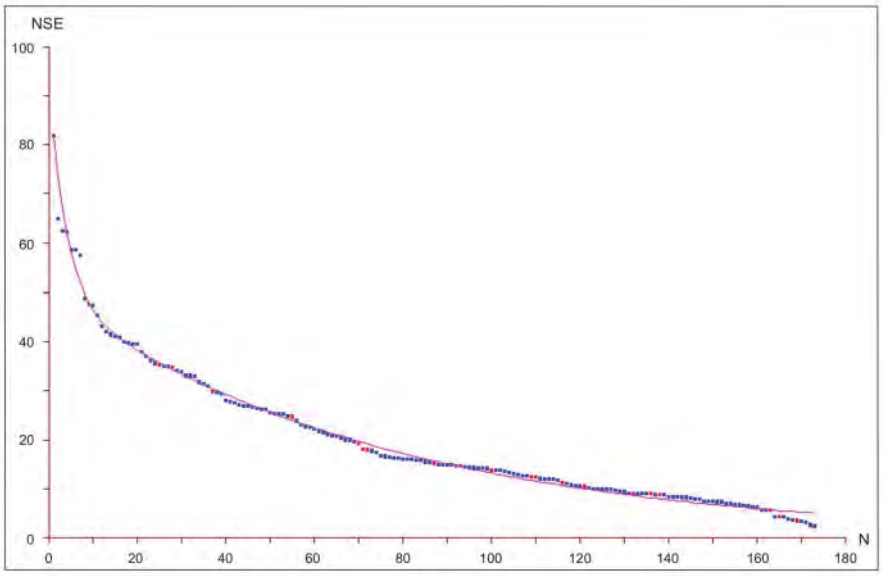

Fig. 4. **Dependence of the normalized squared error (NSE) on the number N of reverse ranking of papillomavirus oncoproteins E6 (173 viruses). Marked with red are E6 oncoproteins of viruses with high oncogenic risk. 2-exponential approximation is:**
$NSE_{HLR} = 49\ EXP\{-0.0132(N-1)\} + 32.852\ EXP\{-0.2639727(N-1)\}$

Assume that d designates the dimension (integer or fractional) of an informational cell that contains NSE parameters of E6 oncoproteins. Then Hartley cumulative entropy is proportional to $N^d LOG(NSE)$. If dimension d is fractional then the informational cell will be fractal; in the case when dimension d is integer, we may suppose that d is the number of linked cells.

Graphics of H^+ и H^- functions of cumulative entropy at $d = 2$ for viruses of high and low oncogenous risk are shown on Figs. 2.2 and 5.

Let us consider the connection of informational cell dimension with features of papillomavirus infection. If $d > 4$, any structural cell changes cannot be observed; that is, the virus infection is in the latent period. It is natural, as multidimentional informational cells just cannot be observed. We may also suppose that dimension d passes all fractional stages from $d = 4$ to $d = 1$.

In the case when $3 < d < 4$, one may observe the productive virus infection for papillomaviruses of high and low oncogenic risk; however for viruses of high oncogenic risk, neoplasia may develop. Virus DNAs are distributed in three-dimensional space of cells for viruses of low oncogenic risk and in four-dimensional space (taking into account the incorporation of virus DNA into the cell genome) for viruses of high oncogenic risk.

Region $2 < d < 3$ defines the fractal dimension of papillomas and condylomas as benign tumors for viruses of low oncogenic risk and epidermoid intraepithelial lesions (precancerous conditions) for papillomaviruses of high oncogenic risk.

Region $d < 2$ for viruses of low oncogenic risk does not exist. Value $d = 2$ may be attributed to pointed condylomas because of their rather soft consistency.

In the region $1 < d < 2$, preinvasive carcinoma is observed when the fractal dimension of cancerous tumour is less then observed two-dimensional one as cancerous tumours are weakly differentiated.

At $d = 1$, metastases appear. Detaching of cancer cells is possible at the cutting of last (single) bond with the cancer tumour.

Before crossing point N_0, we have H– (N) > H+ (N), that is, the entropy of virus DNA elimination or cancer transformation is more than the entropy of quantitative growth of episome DNA.

We will designate the area between curves H– (N) and H+ (N) at $N < N_0$ as S–, and the area between curves H+ (N) and H– (N) at $N > N_0$ — as S+. Then, the probability of cancer cell transformation may be defined as $P_c = S-/S+$.

Table shows probability values of cancer cell transformation for of high and low oncogenic risk.

Note that decreasing rough exponential function NSE (elimination of virus DNA or cancer) corresponds to the return for cell into the past, but not in the least for the cell evolution.

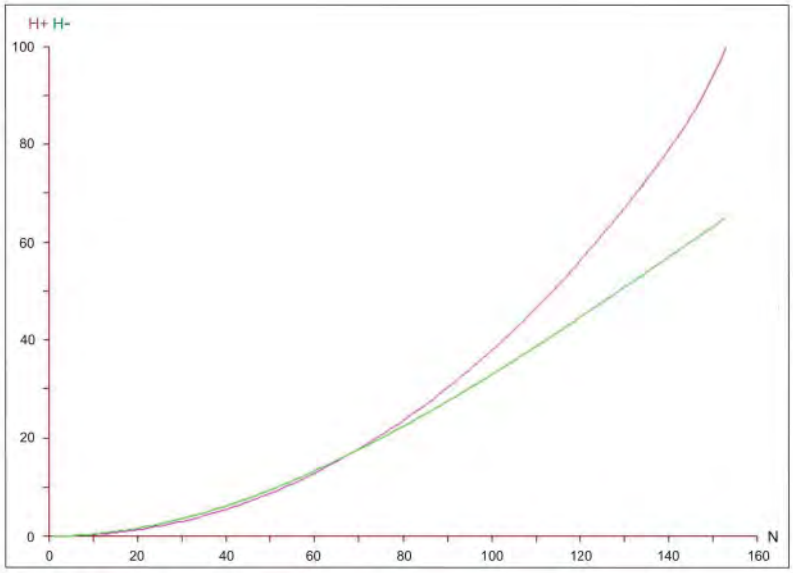

Fig. 5. **Graphics of H^+ и H^- functions of cumulative entropy at $d = 2$ for viruses of low oncogenous risk, N = 1...153:**
$H+ (N) = N^d LOG [50.5 EXP\{- 0.0149(153 - N)\} + 31.352 EXP\{- 0.2812865(153 - N)\}]$.
$H^- (N) = N^d LOG [81.852 EXP\{- 0.02603004(N - 1)\}]$.
Crosspoint of H^+ and H^- is at $N_0 = 67.5$

Values of probability of cancer cell transformation for papillomaviruses of high and low oncogenic risk

d	High risk	Low risk
1	$1.94230300 \cdot 10^{-1}$	—
2	$5.72016772 \cdot 10^{-2}$	$2.92785191 \cdot 10^{-2}$
3	$1.98637659 \cdot 10^{-2}$	$9.06198389 \cdot 10^{-3}$
4	$7.55089981 \cdot 10^{-3}$	$3.06071257 \cdot 10^{-3}$
5	$3.03657242 \cdot 10^{-3}$	$1.09113459 \cdot 10^{-3}$
6	$1.26838686 \cdot 10^{-3}$	$4.03370226 \cdot 10^{-4}$
7	$5.44360273 \cdot 10^{-4}$	$1.53033009 \cdot 10^{-4}$
8	$2.38393159 \cdot 10^{-4}$	$5.91933751 \cdot 10^{-5}$

Evolution for a cell is its transformation at productive virus infection, since at that entropy growth in the cell is the largest.

It is also significant that exponential function NSE for viruses of high oncogenic risk is not too rough. Therefore, external factors heavily influence on cancer cell transformation, as opposed to rough exponent NSE for viruses of low oncogenic risk, when external factors do not influence on cancer cell transformation.

Conclusions

Using a statistical ensemble of E6 proteins separately for viruses of high and low oncogenic risk makes it possible to assess the probability of cancerous transformation of cells, which is proportional to the ratio of the area of entropy of cancer transformation to the area of the productive entropy region papillomavirus infection.

This study did not receive any financial support from a government, community or commercial organization.

The authors state that they have no conflict of interest.

REFERENCES

1. *Walboomers J.M.M., Jacobs M.V., Manos M.M., Bosch F. X., Kummer J. A., Shah K. V., Snijders P.J., Peto J., Meijer C.J., Muñoz N.* Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J. Pathol.* 1999, V. 189, P. 12–19. https://doi.org/10.1002/(SICI)1096-9896(199909)189:1<12::AID-PATH431>3.0.CO;2-F
2. *Harald zur Hausen.* Papillomaviruses in the causation of human cancers – a brief historical account. *Virology.* 2009, V. 384, P. 260–265. https://doi.org/10.1016/j.virol.2008.11.046
3. *Crosbie E. J., Einstein M. H., Franceschi S., Kitchener H. C.* Human papillomavirus and cervical cancer. *Lancet.* 2013, V. 382, P. 889–999. https://doi.org/10.1016/S0140-6736(13)60022-7
4. *Senapati R., Senapati N. N., Dwibedi B.* Molecular mechanisms of HPV mediated neoplastic progression. *Infectious Agents and Cancer.* 2016, V. 11, P. 59. https://doi.org/10.1186/s13027-016-0107-4

5. *Castellsagué X*, *Munoz N*. Chapter 3: Cofactors in human papillomavirus carcinogenesis — role of parity, oral contraceptives and tobacco smoking. *J. Natl. Cancer Inst. Monogr.* 2003, V. 31, P. 20–28.
6. *Castle P. E.* How Does Tobacco Smoke Contribute to Cervical Carcinogenesis? *J. Virol*. 2008, 82 (12), 6084–6086. https://doi.org/10.1128/JVI.00103-08
7. *Long Fu Xi, Hughes J. P., Castle P. E., Edelstein Z. R., Wang C., Galloway D. A.,. Koutsky L. A., Kiviat N. B., Schiffman M.* Viral Load in the Natural History of Human Papillomavirus Type 16 Infection: A Nested Case–control Study. *J. Infect. Dis.* 2011, 203 (10), 1425–1433. https://doi.org/10.1093/infdis/jir049
8. *Yetimalar H., Kasap B., Cukurova K., Yildiz A., Keklik A., Soylu F.* Cofactors in human papillomavirus infection and cervical carcinogenesis. *Arch. Gynecol. Obstet*. 2012, 285 (3), 805–810. https://doi.org/10.1007/s00404-011-2034-3
9. *Jalil E. M., Bastos F. I., dos Santos Melli P. P., Duarte G., Simoes R. T., Yamamoto A. Y., de Morais R. A. A., Quintana S. M.* HPV clearance in postpartum period of HIV positive and negative women: a prospective follow-up study. *BMC Infect. Dis*. 2013, V. 13, P. 564. https://doi.org/10.1186/1471-2334-13-564
10. *Sundstrom K., Ploner A., Dahlstrom L. A., Palmgren J., Dillner J., Adami H.-O., Ylitalo N., Sparén P.* Prospective Study of HPV16 Viral Load and Risk of In Situ and Invasive Squamous Cervical Cancer. *Cancer Epidemiol. Biomarkers Prev*. 2013, 22 (1). 150–158. https://doi.org/10.1158/1055-9965.EPI-12-0953-T
11. *Del Río-Ospina L., Soto-De León S. C., Camargo M., Moreno-Pérez D. A., Sánchez R., Pérez-Prados A., Patarroyo M. E., Patarroyo M. A.* The DNA load of six high-risk human papillomavirus types and its association with cervical lesions. *BMC Cancer*. 2015, V. 15, P. 100. https://doi.org/10.1186/s12885-015-1126-z
12. *Tulay P., Serakinci N.* The role of human papillomaviruses in cancer progression. *J. Cancer Metastasis Treat*. 2016, V. 2, P. 201–213. https://doi.org/10.20517/2394-4722.2015.67
13. *Doorbar J., Quintb W., Banksc L., Bravo I. G., Stoler M., Broker T. R., Stanley M. A.* The Biology and Life-Cycle of Human Papillomaviruses. *Vaccine*. 2012, 30 (5), F55–70. https://doi.org/10.1016/j.vaccine.2012.06.083
14. *Doorbar J., Egawa N., Griffin H., Kranjec C., Murakami I.* Human papillomavirus molecular biology and disease association. *Rev. Med. Virol*. 2015, 25 (1), 2–23. https://doi.org/10.1002/rmv.1822
15. *Egawa N., Doorbar J.* The low-risk papillomaviruses. *Virus Res*. 2017, V. 231, P. 119–127. https://doi.org/10.1016/j.virusres.2016.12.017
16. *Kajitani N., Satsuka A., Kawate A., Sakai H.* Productive Lifecycle of Human Papillomaviruses that Depends Upon Squamous Epithelial Differentiation. *Front Microbiol.* 2012, V. 3, P. 152. https://doi.org/10.3389/fmicb.2012.00152
17. *Akagi K., Li J., Broutian T. R., Xiao W., Jiang B., Rocco J. W., Teknos T. N., Kumar B., Wangsa D., He D., Ried T., Symer D. E., Gillison M. L.* Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res*. 2014, 24 (2), 185–199. https://doi.org/10.1101/gr.164806.113
18. *McBride A. A., Warburton A.* The role of integration in oncogenic progression of HPV- associated cancers. *PLoS Pathog*. 2017, 13 (4), e1006211. https://doi.org/10.1371/journal.ppat.1006211
19. *Pol S. B. V., Klingelhutz A. J.* Papillomavirus E6 oncoproteins. *Virology*. 2013, V. 445, P. 115–137. https://doi.org/10.1016/j.virol.2013.04.026
20. *Roman A., Munger K.* The papillomavirus E7 proteins. *Virology*. 2013, 445 (1–2), 138–168. https://doi.org/10.1016/j.virol.2013.04.013
21. *Maglennon G. A., Doorbar J.* The Biology of Papillomavirus Latency. *Open Virol. J*. 2012, 6 (Suppl 2: M4), 190–197. https://doi.org/10.2174/1874357901206010190
22. *Gravitt P. E.* Evidence and Impact of Human Papillomavirus Latency. *Open Virol. J*. 2012, 6 (Suppl 2: M5), 198–203. https://doi.org/10.2174/1874357901206010198
23. *Brown C. R., Leon M. L., Muñoz K., Fagioni A., Amador L. G., Frain B., Tu W., Qadadri B., Brown D. R.* Human papillomavirus infection and its association with cervical dysplasia in Ecuadorian women attending a private cancer screening clinic. *Braz. J. Med. Biol. Res*. 2009, 42 (7), 629–636. https://doi.org/10.1590/s0100-879x2009000700007
24. *Kaspersen M. D., Larsen P. B., Ingerslev H. J., Fedder J., Petersen G. B., Bonde J., Höllsberg P.* Identification of Multiple HPV Types on Spermatozoa from Human Sperm Donors. *PLoS One*. 2011, 6 (3), e18095. https://doi.org/10.1371/journal.pone.0018095
25. *McBride A. A.* Oncogenic human papillomaviruses. *Phil. Trans. R. Soc. B.* 2017, V. 372, P. 20160273. https://doi.org/10.1098/rstb.2016.0273
26. *Gravitt P. E., Winer R. L.* Natural History of HPV Infection across the Lifespan: Role of Viral Latency. *Viruses*. 2017, 9 (10), 267. https://doi.org/10.3390/v9100267
27. *Plummer M., Schiffman M., Castle P. E.* A 2-Year Prospective Study of Human Papillomavirus Persistence among Women with a Cytological Diagnosis of Atypical Squamous Cells of Undetermined Significance or Low-Grade Squamous Intraepithelial Lesion. *J. Infect. Dis.* 2007, V. 195, P. 1582–1589. https://doi.org/10.1086/516784
28. *Burd E. M.* Human Papillomavirus and Cervical Cancer. *Clin. Microbiol. Rev*. 2003, 16 (1), 1–17. https://doi.org/10.1128/CMR.16.1.1-17.2003

29. *Marks M. A., Castle P. E., Schiffman M., Gravitt P. E.* Evaluation of Any or Type-Specific Persistence of High-Risk Human Papillomavirus for Detecting Cervical Precancer. *J. Clin. Microbiol*. 2012, 50 (2), 300–306. https://doi.org/10.1128/JCM.05979-11
30. *Sudenga S. L., Shrestha S.* Key considerations and current perspectives of epidemiological studies on human papillomavirus persistence, the intermediate phenotype to cervical cancer. *Int. J. Infect. Dis*. 2013, 17 (4), e216–e220. https://doi.org/10.1016/j.ijid.2012.12.027
31. *Schiffman M., Wentzensen N.* Human Papillomavirus Infection and the Multistage Carcinogenesis of Cervical Cancer. *Cancer Epidemiol. Biomarkers Prev*. 2013, 22 (4), 553–560. https://doi.org/10.1158/1055-9965.EPI-12-1406
32. *Santiago D. N., Heidbuechel J. P. W., Kandell W. M.* Fighting Cancer with Mathematics and Viruses. *Viruses*. 2017, 9 (9), 239. https://doi.org/10.3390/v9090239
33. *Pongsumpun P.* Mathematical Model of Cervical Cancer due to Human Papillomavirus Infection. *Mathematical Methods in Science and Engineering. Proceedings of the 1st Int. Conference on Mathematical Methods & Computational Techniques in Science & Engineering (MMCTSE 2014) Athens, Greece – November 28–30, 2014*. P. 157–161.
34. *Ribassin-Majed L., Lounes R.* A SIS model for Human Papillomavirus transmission. MAP5. 2011-03. 15 pages. 2010. https://hal.archives-ouvertes.fr/hal-00555733v1
35. *Obeng-Denteh W., Afrifa R. T., Barnes B. at al.* Modeling the Epidemiology of Human Papilloma Virus Infection and Vaccination and Its Impact on Cervical Cancer in Ghana. *J. Sci. Res. Rep*. 2014, 3 (19), 2501–2518. https://doi.org/10.9734/JSRR/2014/11019
36. *Baussano I., Franceschi S., Plummer M.* Infection transmission and chronic disease models in the study of infection-associated cancers. *Br. J. Cancer*. 2014, V. 110, P. 7–11. https://doi.org/10.1038/bjc.2013.740
37. *Angstmann C. N., Henry B. I., McGann A. V.* A Fractional Order Recovery SIR Model from a Stochastic Process. *Bull. Math. Biol*. 2016, V. 78, P. 468–499. https://doi.org/10.1007/s11538-016-0151-7
38. *Lee S. L., Tameru A. M.* A Mathematical Model of Human Papillomavirus (HPV) in the United States and its Impact on Cervical Cancer. *J. Cancer*. 2012, V. 3, P. 262–268. https://doi.org/10.7150/jca.4161
39. *Dasbach E. J., Elbasha E. H., Insinga R. P.* Mathematical Models for Predicting the Epidemiologic and Economic Impact of Vaccination against Human Papillomavirus Infection and Disease. *Epidemiol. Rev*. 2006, V. 28, P. 88–100. https://doi.org/10.1093/epirev/mxj006
40. *Goldhaber-Fiebert J. D., Stout N. K., Ortendahl J., Kuntz K. M., Goldie S. J., Salomon J. A.* Modeling human papillomavirus and cervical cancer in the United States for analyses of screening and vaccination. *Popul. Health Metr*. 2007, V. 5, P. 11. https://doi.org/10.1186/1478-7954-5-11
41. *Smith R. J., Li J., Mao J., Sahai B.* Using within-host mathematical modeling predict the long-term outcome of human papillomavirus vaccines. *Canadian Applied Mathematics Quarterly*. 2013, 21 (2), 281–299.
42. *Haeussler K., Marcellusi A., Mennini F. S.* Cost-Effectiveness Analysis of Universal Human Papillomavirus Vaccination Using a Dynamic Bayesian Methodology: The BEST II Study. *Value Health*. 2015, 18 (8), 956–968. https://doi.org/10.1016/j.jval.2015.08.010
43. *Myers E. R., McCrory D. C., Nanda K., Bastian L., Matchar D. B.* Mathematical Model for the Natural History of Human Papillomavirus Infection and Cervical Carcinogenesis. *Am. J. Epidemiol*. 2000, 151 (12), 1158–1171. https://doi.org/10.1093/oxfordjournals.aje.a010166
44. *Bureau A., Shiborski S., Hughes J. P.* Applications of continuous time hidden Markov models to the study of misclassified disease outcomes. *Stat. Med*. 2003. 22 (3), 441–462. https://doi.org/10.1002/sim.1270
45. *Kang M., Lagakos S. W.* Statistical methods for panel data from a semi-Markov process, with application to HPV. *Biostatistics*. 2007, 8 (2), 252–264. https://doi.org/10.1093/biostatistics/kxl006
46. *Mitchell C. E., Hudgens M. G., King C. C. at al.* Discrete-time semi-Markov modeling of human papillomavirus Persistence. *Stat. Med*. 2011, 30 (17), 2160–2170.
47. *Spitoni Cristian, Verduijn Marion and Putter Hein*. Estimation and Asymptotic Theory for Transition Probabilities in Markov Renewal Multi–State Models. *Int. J. Biostat*. 2012, 8 (1), 23. https://doi.org/10.1515/1557-4679.1375
48. *Longet S., Schiller J. T., Bobst M., Jichlinski P., Nardelli-Haefliger D.* A Murine Genital-Challenge Model Is a Sensitive Measure of Protective Antibodies against Human Papillomavirus Infection. *J. Virol*. 2011, 85 (24), 13253–13259. https://doi.org/10.1128/JVI.06093-11
49. *De Azambuja K., Barman P., Toyama J.* Validation of an HPV16-mediated Carcinogenesis Mouse Model. *In vivo*. 2014, V. 28, P. 761–768. PMID: 25189887. PMCID: PMC5214601
50. *Hu J., Budgeon L. R., Cladel N. M., Balogh K., Myers R., Cooper T. K., Christensen N. D.* Tracking vaginal, anal and oral infection in a mouse papillomavirus infection model. *J. Gen. Virol*. 2015, 96 (12), 3554–3565. https://doi.org/10.1099/jgv.0.000295
51. *Jiafen Hu, Nancy M. Cladel, Lynn R. Budgeon, Balogh K. K., Christensen N. D.* The Mouse Papillomavirus Infection Model.

Viruses. 2017, V. 9, P. 246. https://doi.org/10.3390/v9090246
52. *Mandelbrot B.* An information theory of the statistical structure of language. In *Communication Theory. ed. W. Jackson, Betterworths*. 1953, P.486–502.
53. *Hartley R. V. L.* Transmission of information. *Bell Syst. Techn. J.* 1928, 7 (3), 535–563. https://doi.org/10.1002/j.1538-7305.1928.tb01236.x

ВИЗНАЧЕННЯ ЙМОВІРНОСТІ РАКОВОЇ ТРАНСФОРМАЦІЇ КЛІТИН ЗА ПАПІЛОМАВІРУСНОЇ ІНФЕКЦІЇ ЛЮДИНИ

Л. П. Бучацький
В. В. Щербик

Київський національний університет імені Тараса Шевченка, Україна

E-mail: iridolpb@gmail.com

Мета. Оцінити ймовірності ракової трансформації клітин вірусами високого та низького онкогенних ризиків.

Методи. Використовували статистичні ансамблі визначення ймовірності трансформації ракових клітин.

Для аналізу було відобрано онкопротеїни E6, які безпосередньо впливають на процес трансформації ракових клітин.

Результати. За допомогою нормальної квадратичної помилки (NSE) для вірусів високого (20 штамів) і низького (153 штами) онкогенних ризиків було побудовано рангову статистику 2-експоненціального типу. Для продуктивної папіломавірусної інфекції функцію NSE визначали як зростаючу точну 2-експоненту клітинного шару базальної поверхні епітелію. Логарифм числових значень NSE пропорційний ентропії клітин, яка пов'язана з наявністю вірусної ДНК. Для підрахунку ентропії використовували узагальнену формулу Хартлі з інформаційною коміркою розмірності d: H = NdLOG (NSE), де N — узагальнена координата комірки.

Висновки. Використання статистичного ансамблю протеїнів E6 окремо для вірусів високого та низького онкогенних ризиків дає змогу оцінити ймовірність ракової трансформації клітин, яка пропорційна відношенню площі ентропії ракової трансформації до площі продуктивної ентропії папіломавірусної інфекції.

Ключові слова: папіломавірусна інфекція людини, канцерогенне переродження клітин, кумулятивна ентропія Хартлі.

ОПРЕДЕЛЕНИЕ ВЕРОЯТНОСТИ РАКОВОЙ ТРАНСФОРМАЦИИ КЛЕТОК ПРИ ПАПИЛЛОМАВИРУСНОЙ ИНФЕКЦИИ ЧЕЛОВЕКА

Л. П. Бучацкий
В. В. Щербик

Киевский национальный университет имени Тараса Шевченко, Украина

E-mail: iridolpb@gmail.com

Цель. Оценка вероятности раковой трансформации клеток вирусами высокого и низкого онкогенных рисков.

Методы. Использовались статистические ансамбли для определения вероятности трансформации раковых клеток.

Для анализа были отобраны онкопротеины E6, непосредственно влияющие на процесс трансформации раковых клеток.

Результаты. С помощью нормализованной квадратичной ошибки (NSE) для вирусов высокого (20 штаммов) и низкого (153 штаммов) онкогенного риска была построена ранговая статистика 2-экспоненциального типа. Для продуктивной папилломавирусной инфекции функцию NSE определяли как возрастающую точную 2-экспоненту клеточного слоя базальной поверхности эпителия. Логарифм числовых значений NSE пропорционален энтропии клетки, связанной с наличием вирусной ДНК. Для вычисления энтропии использовалась обобщенная формула Хартли с информационной ячейкой размерности d: H = NdLOG (NSE), где N — обобщенная координата ячейки.

Выводы. Использование статистического ансамбля протеинов E6 отдельно для вирусов высокого и низкого онкогенных рисков позволяет оценить вероятность раковой трансформации клеток, которая пропорциональна отношению площади энтропии раковой трансформации к площади продуктивной энтропии папилломавирусной инфекции.

Ключевые слова: папилломавирусная инфекция человека, канцерогенное перерождение клеток, кумулятивная энтропия Хартли.