ORIGINAL RESEARCH

Analysis of *Salmonella typhimurium* Protein-Targeting in the Nucleus of Host Cells and the Implications in Colon Cancer: An in-silico Approach

Jianhua Li [ID] [1]
Mohammed Zakariah [2]
Abdul Malik [3]
Mohammad Shamsul Ola [ID] [4]
Rabbani Syed [3]
Anis Ahmad Chaudhary [ID] [5]
Shahanavaj Khan [ID] [6,7]

[1]Department of General Surgery Ⅰ, Xinxiang Central Hospital, Xinxiang City, Henan Province 453000, People's Republic of China; [2]Research Center, College of Computer and Information Science, King Saud University, Riyadh, Saudi Arabia; [3]Department of Pharmaceutics, College of Pharmacy, King Saud University, Riyadh 11451, Saudi Arabia; [4]Department of Biochemistry, College of Science, King Saud University, Riyadh, Saudi Arabia; [5]Department of Biology, College of Science, Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia; [6]Bioinformatics and Biotechnology Unit, Department of Bioscience, Shri Ram Group of College (SRGC), Muzaffarnagar, UP, India; [7]Nano-Biotechnology Unit, Department of Pharmaceutics, College of Pharmacy, King Saud University, Riyadh 11451, Saudi Arabia

Correspondence: Shahanavaj Khan
Tel +91 9219993262
Email khan.shahanavaj@gmail.com

Background: Infections of *Salmonella typhimurium* (*S. typhimurium*) are major threats to health, threats include diarrhoea, fever, acute intestinal inflammation, and cancer. Nevertheless, little information is available about the involvement of *S. typhimurium* in colon cancer etiology.

Methods: The present study was designed to predict nuclear targeting of *S. typhimurium* proteins in the host cell through computational tools, including nuclear localization signal (NLS) mapper, Balanced Subcellular Localization predictor (BaCeILo), and Hum-mPLoc using next-generation sequencing data.

Results: Several gene expression-associated proteins of *S. typhimurium* have been predicted to target the host nucleus during intracellular infections. Nuclear targeting of *S. typhimurium* proteins can lead to competitive interactions between the host and pathogen proteins with similar cellular substrates, and it may have a possible involvement in colon cancer growth. Our results suggested that *S. typhimurium* releases its proteins within compartments of the host cell, where they act as a component of the host cell proteome. Protein targeting is possibly involved in colon cancer etiology during intracellular bacterial infection.

Conclusion: The results of current in-silico study showed the potential involvement of *S. typhimurium* infection with alteration in normal functioning of host cell which act as possible factor to connect with the growth and development of colon cancer.

Keywords: *S. typhimurium*, in- silico analysis, proteome, nuclear targeting protein, colon cancer etiology

Introduction

About 20% of the overall burden of cancer can presently be associated with infections of different types of agents, such as viruses, parasites and bacteria.[12,58] A largely neglected aspect of infectious diseases is the connection of chronic bacterial infections with the growth and development of cancer,[4] among very few bacterial infections associated with cancer development to date[2,3,47] with respect to viral-mediated cancers.[7,13] Although studies have shown that specific strains of bacteria are associated with the growth of different types of human cancers but their molecular mechanism is not well understood. Previous studies have connected specific bacterial species to the carcinogenesis of different types of cancers, however, various bacterial species have emerged as therapeutic agents in the prevention, diagnosis, or management of cancers.[34]

Different bacterial strains may be connected with the growth and development of different types of cancers by stimulating DNA damaging processes through toxins, inflammation and alteration in metabolites, and normal cell signaling pathways of the host during infections.[6,53] For example, pathogenic strains of *E. coli* have the ability to change normal functions of the host cell through induction of chronic inflammation and interference with the host cell cycle, recommending a potential connection between specific bacteria and cancers.[18,22] Similarly, *Helicobacter pylori* has been connected as a possible risk factor for the growth of various types of stomach cancers.[43] Bacteria has the ability to change normal functions of the host cell during infection, as demonstrated by pathogenic species of *Salmonella*, which manipulate signaling pathways of the host cell for intracellular survival and bacterial uptake.[11]

If specific bacterial strains can induce cancer development by altering the cell signaling in the host,[10] it is possible that infections of specific bacterial strains would increase the risk of cancer. In particular, this could occur in cases of long-term bacterial infections, where the possibility of targeting a formerly pretransformed cell is higher. The effector proteins of *Salmonella* activate the AKT/ERK signalling pathway in host cells. It has been observed that the AKT/ERK pathway is also stimulated in different types of cancers.[48] The effector AvrA of *Salmonella* stimulates β-catenin signaling in the infected host cell, which supports carcinogenesis in the colon of mice[32,33] Although these studies demonstrated the association of bacterial infection in the growth of colon and colorectal cancers in human clinical specimens and experimental mouse models,[32] it remains uncertain whether *Salmonella* infections acts as a possible cause for colon cancer in humans. The frequency of colon cancer increases over time through various unidentified potential factors.[5,56]

To determine whether *Salmonella* infections represent another possible factor for colon cancer development, we predicted the nuclear targeting of *Salmonella typhimurium* strain Ty proteins in the host cell using next-generation sequencing data of whole proteome from the UniProt database. Moreover, we examined the implication of such nuclear targeting proteins in the etiology of colon cancer during *Salmonella* infection.

Methods
Selection of *Salmonella typhimurium* Proteome
The database of Universal Protein Resource (UniProt) was utilized for the selection of specific strains of *S. typhimurium* to analyze nuclear protein targeting in the current work.[1] The whole *S. typhimurium* proteome was retrieved from UniProt and utilized to predict nuclear protein targeting in host cells and their involvement in colon cancer development.[36,45]

Bioinformatics Tools for Prediction of *Salmonella typhimurium* Nuclear Targeting Proteins in Host Cells
S. typhimurium LT2 strain whole proteome was selected for computation of nuclear targeting proteins in the host cell by employing cNLS mapper, Balanced Subcellular Localization (BaCelLo), and Hum-mPLoc 2.0 bioinformatics tools.[45]

cNLS Mapper for Prediction of Nuclear Localization Signals in *Salmonella typhimurium* Proteins
The whole proteome of *S. typhimurium* LT2 was utilized to predict the nuclear localization signal (NLS) using the bioinformatics tool cNLS mapper.[26] The cNLS mapper generated activity-based reports for diverse categories of importin-α-dependent NLSs, which characterize the functional roles of diverse amino acids at each position within an NLS class. *S. typhimurium* protein sequences were predicted as follows: particularly targeted to the cytoplasm, targeted to the cytoplasm as well as the nucleus, partially targeted to the nucleus, and particularly targeted to the nucleus with a specific range of cutoff values of 1–2, 3–5, 7–8, and 8–10, respectively, as demonstrated in the previous cNLS literature.[26]

BaCelLo Predictor for Prediction of *Salmonella typhimurium* Nuclear Localization Proteins
Nuclear targeting proteins of *S. typhimurium* LT2 were predicted using the Balanced Subcellular Localization (BaCelLo) tool. The BaCelLo predictor is an important bioinformatics software for the prediction of protein localization in the eukaryotic cell. It is worked on diverse support vector machines (SVMs) that can predict subcellular protein targeting in five different organelles of eukaryotes, such as

the nucleus, mitochondrion, cytoplasm, plasma membrane (secretory proteins) and chloroplast.[42]

Hum-mPLoc 3.0 Predictor for Prediction of *Salmonella typhimurium* Nuclear Localization Proteins

The Hum-mPLoc 2.0 predictor was employed to confirm nuclear protein targeting in humans using whole proteins from the *S. typhimurium* LT2 proteome. The bioinformatics tool Hum-mPLoc 2.0 operates on a top-down networking system.[50] The bioinformatics Hum-mPLoc 2.0 predictor can predict protein targeting in 14 different compartments of the cell, including the cytoplasm, mitochondria, endoplasmic reticulum, centrioles, Golgi apparatus, and nucleus.

Results
Selection of *Salmonella typhimurium* Proteome

The UniProt database is a widespread source for protein sequences, which was developed through the collection of Swiss-Prot, PIR and TrEMBL protein database information.[1] This is a freely accessible comprehensive resource and database for data annotation and protein sequence.[44] The proteome of *S. typhimurium* LT2 strain was chosen to analyze nuclear targeting proteins in the host cell, which contained the highest number of proteins sequences,[45] with respect to all existing strains.[36]

Bioinformatics Tools for Prediction of *Salmonella typhimurium* Nuclear Targeting Proteins in Host Cells

Different bioinformatic tools were used in the current study, such as cNLS mapper, Balanced Subcellular Localization (BaCelLo), and Hum-mPLoc 2.0 software, for the analysis of *S. typhimurium* nuclear targeting proteins. These predictors were selected for the prediction of nuclear targeting proteins.

cNLS Mapper for Prediction of Nuclear Localization Signals in *Salmonella typhimurium* Proteins

It was observed that increase in molecular weight constantly enhanced nuclear targeting, except for the range of 20–40 kDa. The 20–40 kDa molecular weight proteins were targeted primarily in the nucleus of the host cell. Moreover, the

isoelectric point (pI) value did not demonstrate specific pattern for nuclear targeting. The patterns of *S. typhimurium* proteins targeting in the host cell nucleus using different parameters are described in Figure 1. The whole proteins targeting of *S. typhimurium* in various components of host cells with diverse parameters is illustrated in graph (Figure 2). Supplementary Table S1 shows details characteristics of nuclear targeted proteins of host cell.

Briefly, 141 nuclear targeted proteins of *S. typhimurium* were predicted with different values of monopartite and bipartite NLSs. Among those, 121, 9, 10, and 1 proteins were found to have 0–3.0, 3.0–5.0, 5.0–8.0, and >8.0 cutoff values for monopartite NLSs, respectively. Furthermore, among the 141 nuclear targeted proteins, 32, 84, 23, and 2 proteins were found to have 0–3.0, 3.0–5.0, 5.0–8.0, and >8.0 cutoff values for bipartite NLSs, respectively (Figure 1). Increasing the cutoff value of monopartite NLS is connected with declined nuclear targeting, and a similar pattern was found with bipartite NLS values, except for cutoff values in the range of 3.0–5.0 (Figure 1). The detailed results of nuclear targeting proteins after synchronization are demonstrated in Supplementary Table S1.

Similarly, whole protein sequences (4718) of *S. typhimurium* were predicted with different values of monopartite and bipartite NLSs. Among those, 4356, 173, 135, and 54 proteins were found to have 0–3.0, 3.0–5.0, 5.0–8.0, and >8.0 cutoff values for monopartite NLSs, respectively. Furthermore, among 4718 proteins, 1635, 2564, 499, and 20 proteins were found to have 0–3.0, 3.0–5.0, 5.0–8.0, and >8.0 cutoff values for bipartite NLSs, respectively (Figure 2).

BaCelLo Predictor for Prediction of *Salmonella typhimurium* Nuclear Localization Proteins

The outcomes of BaCelLo showed that 331 *S. typhimurium* proteins were targeted to the host cell nucleus. The results of prediction analysis also demonstrated that 2069 proteins were targeted in the cytoplasm, 921 in the mitochondria, and 1397 were secretory proteins (Table 1).

Hum-mPLoc 2.0 Predictor for Prediction of *Salmonella typhimurium* Nuclear Localization Proteins

To increase the prediction efficiency, we used the Hum-mPLoc bioinformatics tool, which operates using a top-down strategy to predict protein targeting in different
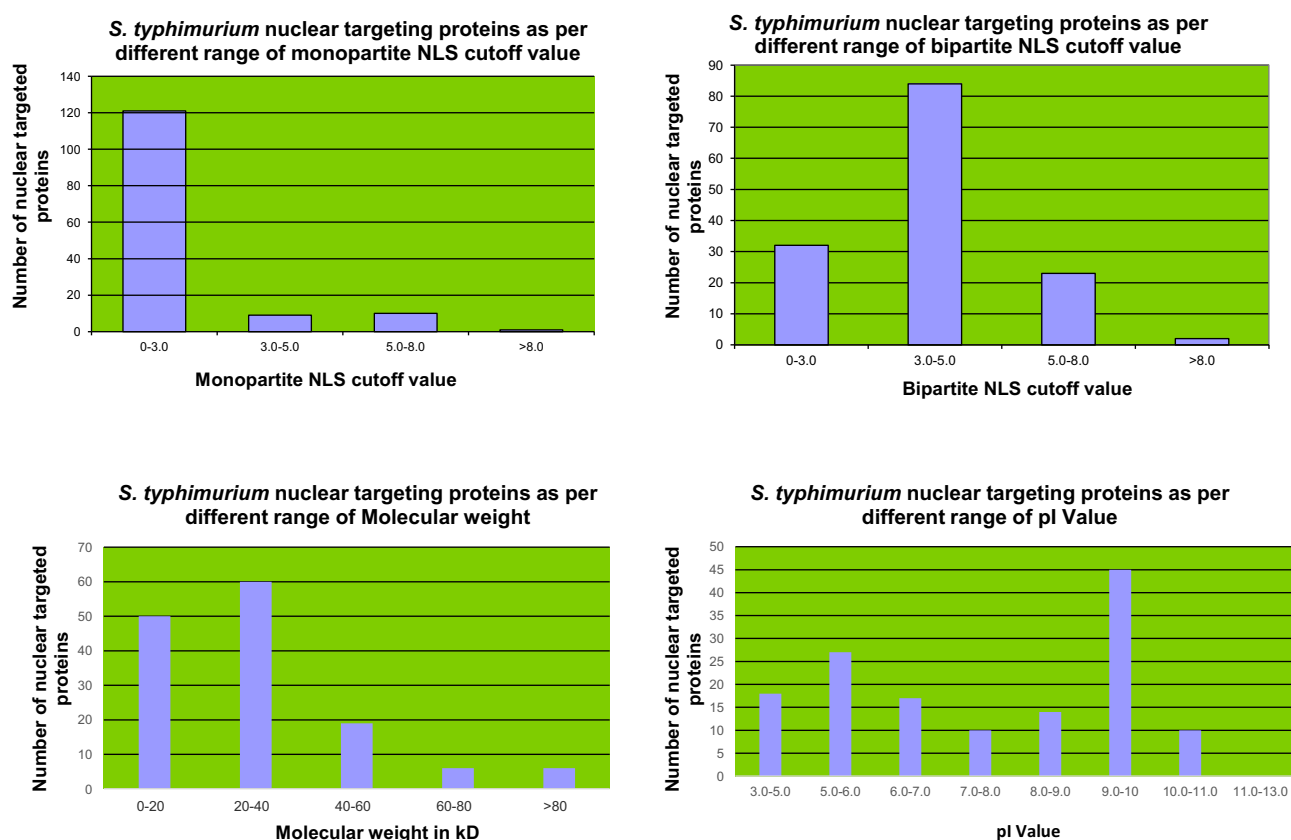
S. typhimurium nuclear targeting proteins as per different range of monopartite NLS cutoff value

S. typhimurium nuclear targeting proteins as per different range of bipartite NLS cutoff value

S. typhimurium nuclear targeting proteins as per different range of Molecular weight

S. typhimurium nuclear targeting proteins as per different range of pI Value

Figure 1 Computational prediction of nuclear targeting of *S. typhimurium* proteins in host cells and their relationship with various parameters.

components of human cells, including the nucleus.[49] The results of *S. typhimurium* protein targeting illustrated exclusive targeting of proteins in the following regions using the Hum-mPLoc tool: 867 in the nucleus, 707 in the cytoplasm, 654 in the mitochondria, 289 in the plasma membrane, 215 in the endoplasmic reticulum, 15 in the Golgi apparatus, 64 in peroxisomes, 1 in microsomes, 18 in the lysosome, extracellular 444, 1 in the centrosome, and 83 in unknown locations. The prediction results showed 1360 targeted proteins in multiple organelles (Table 1).

Synchronization of BaCelLo and Hum-mPLoc 2.0 Results

The outcome of synchronization showed that among BaCeILo predicted total nuclear (331), mitochondrial (921), secretory (1397), and cytoplasmic (2069) proteins. When the BaCeILo predicted proteins were compared with the Hum-mPLoc 2.0 results, only 141 proteins showed consistent results with respect to nuclear targeting proteins.

Detailed results and properties of nuclear targeting proteins are illustrated in Supplementary Table S1. Recent evidences

showed the connection of *Salmonella* infection with the development of colorectal cancer in human.[38] It is observed that the effector AvrA of *Salmonella* activates host β-catenin signaling pathways which promotes the carcinogenesis in colon of mice.[32,33] Our results showed the targeting of bacterial *S. typhimurium* DNA mismatch repair protein MutS (Accession no. P0A1Y1) in the nucleus of host cells. DNA mismatch repair proteins (MMRPs) are ubiquitous performers in a wide range of main cellular functions.[16] The rate of spontaneous mutation increased due to alteration in the MMR, which affects the growth of different types of cancer, including colon cancer.[17] A previous reportshowed the possible role of *Escherichia coli* DNA MMRPs in the growth and development of colon cancer.[22] In this study, the results illustrated nuclear targeting of various DNA binding proteins, such as DNA directed RNA polymerase subunit alpha (Accession No. P0A7Z8), DNA polymerase III subunit epsilon (Accession No. P0A1H0), replication protein RepA (Accession No. Q934T6), putative DNA-binding protein (Accession No. Q934Z5, Q8Z752), DNA-invertase (Accession No. Q8Z339), RNA polymerase-binding transcription factor DksA (Accession No. P0A1G6), RNA polymerase
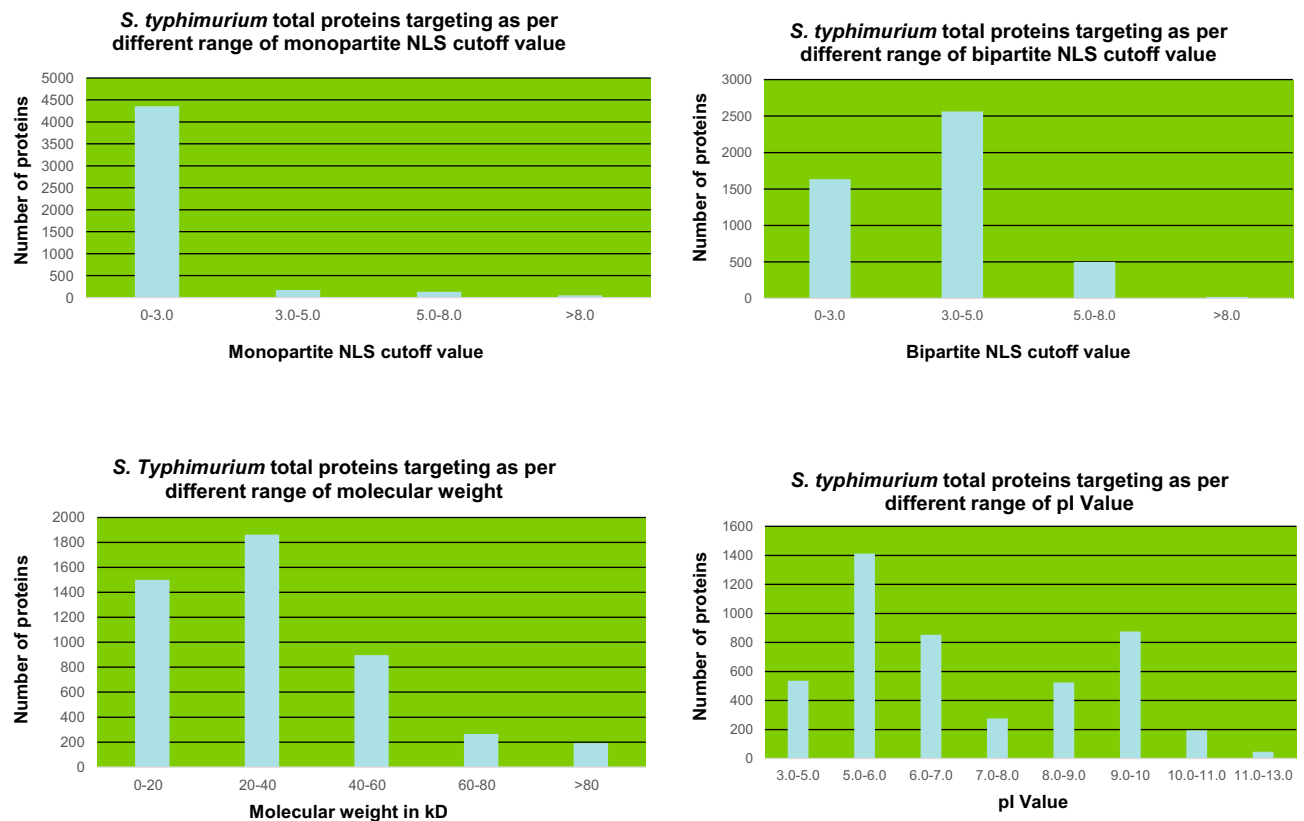
S. typhimurium total proteins targeting as per different range of monopartite NLS cutoff value

S. typhimurium total proteins targeting as per different range of bipartite NLS cutoff value

S. Typhimurium total proteins targeting as per different range of molecular weight

S. typhimurium total proteins targeting as per different range of pl Value

Figure 2 Computational prediction of total targeting of S. typhimurium proteins in host cells and their relationship with various parameters.

sigma factor RpoD (Accession No. P0A2E4), ATP-dependent RNA helicase (Accession No. Q8Z877), DNA topoisomerase III (Accession No. Q8Z6F5), topoisomerase IV subunit A (Accession No. Q8Z3P4), topoisomerase B (Accession No. Q9RHF5), putative DNA-binding protein (Accession No. Q8Z2J9), excinuclease cho (Accession No. Q8Z6G5), and RepHI1A Replication initiation protein (Accession No. Q7BRX0).

Moreover, the results show that RNA chaperone ProQ (Accession No. P60318), chaperone protein SigE (Accession No. Q8Z7R2), chaperone protein DnaJ (Accession No. P0A1G8), and chaperone modulatory protein CbpM (Accession No. Q8XGQ8) were localized in the nucleus of the host cell. The localization of these proteins in the nucleus of the host cell suggests their potential implications in colon cancer growth, but this requires further research.

Discussion

Protein targeting of various microbes in specific host cells has immense impact on regulation of different pathways of host cells, which may promote normal-to-cancer cell transformation. Previously, published report has showed that M. hominis proteins targeting in the nucleus of host cells

and their possible involvement in growth of prostate cancer.[24] In thisstai study we have selected BaCeILo predictor, which utilizes the data set of different animals, fungi and plants. We have also selected another predictor, Hum-mPLoc, for its high accuracy in our results. Although the possible implication of S. typhimurium on the growth and development of colon cancer has been previously studied, the exact mechanism is not well understood. Also few studies showed that the attenuated strain (VP20009) of S. typhimurium emerges to have a nontoxic profile in humans, while it has been utilized in Phase I trials for the management of colon cancer and melanoma patients with less toxic effects. Mutations have been connected with the lesser toxicity of this strain, which decreases the toxicity of its lipopolysaccharide.[9,39]

A recent study has demonstrated that severe salmonellosis has connected with the high risk of development of colon cancer in the ascending or transverse parts of the colon.[38] The bacterial proteins or effectors play important role in altering normal functioning of different pathways of host cell. For instance, the AvrA protein of Salmonella is targeted to the host cells through Type Three Secretion System (TTSS), which suppresses the process of program cell death/

Table 1 Details of Possible Protein Targeting in Different Subcellular Locations of the Host Cell by BaCelLo and Hum-mPLoc 2.0, Using the Complete *Salmonella typhimurium* Proteome

Protein Targeting in Different Subcellular Locations	Number of Proteins Predicted by BaCelLo	Number of Proteins Predicted by Hum-mPLoc 2.0
Nucleus	331	867
Mitochondria	921	654
Cytoplasm	2069	707
Secretory/Plasma membrane	1397	–
Plasma Membrane	–	289
Endoplasmic reticulum	–	215
Golgi apparatus	–	15
Peroxisome	–	64
Microsome	–	01
Lysosome	–	18
Extracellular	–	444
Centrosome	–	01
Cytoskeleton	–	0
Multiple location	–	1360
Unknown	–	83

apoptosis, specifically with reference to enteropathogenic salmonellosis for prolonging intracellular bacterial survival.[31] Various bioinformatics predictors for prediction of subcellular targeting of object proteins are available, which mainly rely on prediction of specific localization motifs on proteins or search for similarity/alignment to predict possible localization of similar proteins in new systems.[14,46] The analysis of NLSs in particular proteins is very important for predicting their nuclear localization.[51] All proteins generally carry a specific sequence motif known as the NLS. There are 6 classes of NLSs that have been detected, which are involved in nuclear import of proteins through α/β importin pathways. Two types of NLSs have been categorized based on stretches of the basic amino acids, i.e., monopartite (1 basic stretch) and bipartite (2 basic stretches). These 2 stretches mediate the binding of NLS to the importin-α transport receptor. Such complex attaches to importin-β, stimulating localization of a specific protein to the nucleus.[29] Our rationale behind using the cNLS mapper was to predict the activity of NLS instead of an NLS sequence. The cNLS mapper identifies the involvement of each residue in the NLS and analyzes NLS activity which is proposed to offer highly accurate analysis performance.[25,26] However, the cNLS mapper is unable to predict any proteins that directly bind to importin-β.

Moreover, the cNLS mapper predicts NLS activity as an isolated peptide, but not within the overall structural context of the protein.

Another subcellular localization prediction tool, BaCeILo, was used in our study, and it relies on different SVMs to predict secretory, nuclear, cytoplasmic, chloroplast, and mitochondrial targeting of certain proteins.[42] It predicts particular targeting based on alignment to ascertain evolutionary information and residue sequence information within the whole protein sequence and its N and C termini.

Additionally, we have also used Hum-mPLoc, which is based on the human protein subcellular location prediction approach to predict subcellular location using the information of peptide composition, amino acid composition, and similarity.[49] Hum-mPLoc also detects protein localization based on a hybrid method using all three approaches. The results of the localization also provide more reliability and accuracy in order to estimate the certainty of the prediction results. We used a hybrid approach as, according to the literature, it has been shown to give better accuracy.

Our aim to obtain prediction results from diverse methods was satisfied by the use of these tools as all these tools work in different ways. The NLS mapper works using data generated from yeast and therefore its accuracy to predict NLS in humans can be evaluated. Both BaCeILo and Hum-mPLoc 2.0 predictors do not show the same pattern of protein targeting due to the use of different algorithms and data sets by both computational tools. The BaCelLo predictor is based on diverse SVMs organized in a decision tree and utilizes the information obtained from the residue sequence and from the evolutionary information contained in alignment profiles of different animals, fungi, and plants. The predictor has utilized variable numbers of sequences for nucleus (Plants: 121, Animals: 1166, Fungi: 711), cytoplasm (Plants: 58, Animals: 439, Fungi: 211), extracellular (Plants: 41, Animals: 804, Fungi: 88), mitochondria (Plants: 67, Animals: 188, Fungi: 188), and chloroplast (Plants: 204). It analyzes the whole sequence composition and the compositions of both the N- and C-termini. Although the training set is fully curated in order to avoid any possible redundancy during prediction but used information of three kingdom systems. For the first time a balancing procedure is introduced in order to diminish the cause of biased training sets.

Analyzing subcellular localization in human proteins is a very challenging task due to the existence of multiplex character in query proteins. On the other hand, the Hum-mPLoc utilizes human specific protein data sets during analysis of protein sequences. The Hum-mPLoc predictor worked

on learning data set includes 3681 protein sequences (3106 different proteins), classified into 14 human subcellular locations including nucleus, cytoplasm, cytoskeleton, endoplasmic reticulum, centriole, endosome, extracellular, Golgi apparatus, lysosome, microsome, mitochondrion, plasma membrane, synapse and peroxisome. The Hum-mPLoc 2.0 predictor worked on a top-down strategy using sequences information from human data set exclusively. The Hum-mPLoc 2.0 predictor utilized the sequential evolution information and functional domain information as ensemble classifier. Therefore, the results of BaCeILo were used to narrow down proteins as per their localization prediction in the animal-specific predictor, and they were further scrutinized after using Hum-mPLoc 2.0 tools, which has used the human-specific data set. It has been estimated that only 30% of nuclear proteins actually carry NLS sequences; therefore, a large proportion of protein can localize to the nucleus, even in the absence of NLS.[8,26] In addition, small proteins with less than 40 kDa molecular weights can passively diffuse into the nucleus.[15,23] The methods used for protein targeting predictions are different with every tool therefore, variation in results is logical.

Nuclear Targeting Proteins of *S. typhimurium* and Their Role in Cancer

Nuclear targeting proteins of *S. typhimurium* can have a variety of consequences on progression of colon cancer etiology. Additional reports discuss the potential of nuclear-targeted proteins in cervical cancer etiology. The Hum-mPLoc operates by using an experimentally annotated data set of 3780 human proteins. Although it is designed to predict subcellular localization of human proteins, its performance in predicting protein subcellular localization of several closely related mammalian genomes is also plausible. However, its role in predicting subcellular localization of bacterial proteins as the query is arguable and the results need to be experimentally validated. Furthermore, both BaCeILo and Hum-mPLoc are based on SVM with different data sets. SVMs have certain limitations, especially when a different data set is used for analysis. Therefore, the differences in results obtained from different tools can be understood.

DNA Binding Proteins

The role of DNA binding proteins is very important in cancer development. For example chromodomain helicase DNA binding protein 5 is involved in tumor suppression, and mutation in the protein synthesis can lead to loss of its

function and further development of breast cancer in humans.[57] Certain DNA binding proteins, like CpG binding proteins, are involved in cancer progression through methylation of target DNA.[41] Nucleotide excision repair protein DDB2 (DNA damage binding protein 2) is another DNA binding protein that is suggested to cause cancer after loss of its activity.[20] The existence of both *S. typhimurium* and human DNA binding proteins in the host nucleus provides competition for both proteins to bind their substrates. Similar substrates can interfere with the binding of normal human proteins to the target site and can affect progression of colon cancer. The role of DNA binding inhibitor proteins in cancer development has already been proven in ovarian cancer, where inhibition of DNA binding protein (ID-1) overexpression leads to ovarian cancer.[35]

DNA Repair Proteins

Erroneous DNA repair has been implicated as an etiological factor in several cancers. For example, the MutS protein, which is involved in DNA mismatch repair, is involved in the colon cancer etiology.[22,40] In this study we found nuclear localization of DNA MMRP MutS (Accession No. P0A1Y1), Host-nuclease inhibitor protein (Accession No. Q8Z7Y2), and Putative DNA replication terminus site-binding protein (Accession No. Q9L5F9). Alteration in mismatch repair genes is already linked with cancer development, and human cells have a homolog for MutS. MMR mutation enhanced the rates of spontaneous mutation which may act as a potential factor for the development of cancer including colon cancer.[17] Recent research work reported the implication of *S. typhimurium* in growth of colon cancer through down regulation of Wnt1 using a Salmonella-colitis colon cancer model.[55] In an early report, it has shown the possible role of *Escherichia coli* DNA MMRPs including Muts in growth and development of colon cancer.[22] Therefore, it can be supposed that if two homologous proteins with the same enzymatic function are present in a particular cell, the relative enzymatic function of proteins will be different, and both of them compete with each other to bind with their substrate. As MutS is a DNA repair associated protein, aberration in DNA repair can lead to the development of cancer. It has already been proposed that it induces DNA damage in host cells and suppresses DNA repair activity, but this suppressed DNA repair involves double-stranded breaks.[38] Future research on the involvement of *S. typhimurium* nuclear targeted proteins in suppression of mismatch DNA repair activity can provide new

insights on the role of *S. typhimurium* in the development of colon cancer.

DNA Damage Proteins

In addition, certain DNA damaging endonucleases are involved in the development of cancer. For example, a variant LINE-1 endonuclease overexpression is involved in the development of primary gastric cancer and lymph node metastasis.[54] Our prediction results demonstrated nuclear localization of *S. typhimurium* ribonuclease R (Accession No. Q8XF68) and endoribonuclease SymE (Accession No. Q8Z0W8). In addition to DNA damaging activity, ribonuclease R is predicted to localize to the host cell nucleus and its role in colon cancer etiology needs to be investigated further.

Transcriptional and Translational Regulators

It has been found that RNA polymerase subunits share structural similarities and antigenicity among eukaryotes. This indicates that RNA polymerase is somewhat conserved. Yeast RNA polymerase II has subunit RBP, which binds to nucleotides, and shares similarities with a β subunit of *E. coli* RNA polymerase.[52] Various transcription-associated proteins have the ability to bind DNA; therefore, they can also inhibit binding of host gene regulators and modulate host gene expression.[19] Many transcription factors are involved in cancer development.[30]

The results of the current study indicate that translation initiation factor IF-1 (Accession No. P69225), tRNA (Met) cytidine acetyltransferase TmcA (Accession No. Q8Z4S1), ATP-dependent RNA helicase (Accession No. Q8Z3I1), ATP-dependent RNA helicase RhlB (Accession No. P0A2P1), dual-specificity RNA methyltransferase (Accession No. Q8Z4P2), 23S rRNA pseudouridine synthase (Accession No. Q8Z1V3), putative transcriptional regulator (Accession No. Q8Z6E4), putative reverse transcriptase (Accession No. Q8Z4H7), transcriptional regulatory protein BtsR (Accession No. Q8Z5C1), LysR-family transcriptional regulator (Accession No. Q8Z3S5), ribosome-recycling (RRF) factor/ribosome-releasing factor (Accession No. P66739), transcriptional activator CaiF (Accession No. Q8XFZ2), transcriptional activator RamA (Q8Z8M2), met repressor (Accession No. Q8Z2Z4), ribosomal protein S12 methylthiotransferase RimO (Accession No. Q8Z861), putative RNA methyltransferase (Accession No. Q8XFT1), ribosomal large subunit pseudouridine synthase B (Accession No. Q8Z7D5), and transcriptional regulatory protein OmpR (Accession No. P0AA20) localize to the host nucleus. The nuclear localization of transcription associated *S. typhimurium* proteins should be investigated further for their potential involvement in the regulation of human gene expression. The role of bacterial RNA polymerase in human gene transcription has already been investigated in many cases. The RNA polymerase II of *E. coli* has been shown to transcribe human DNA, known as the arrest site, which indicates that bacterial transcriptional regulators can also act on human DNA sequences.[37]

In addition, the variation in human and *S. typhimurium* promoter sequences raises criticism for practical application of *S. typhimurium* transcriptional proteins in human gene transcription. Therefore, it can be anticipated that computational prediction has its own limitations, and the data obtained needs to be validated experimentally before the role of *S. typhimurium* transcription-associated proteins in human genes transcription can be concluded. Presently, it can be assumed that the localization of such proteins can cause several alterations in human gene expression and may possibly contribute to *S. typhimurium*-related colon cancer etiology.

Lateral Gene Transfer and RNA Chaperone

Recently, it has been discovered that lateral gene transfer is a common event during chronic infections, where bacteriato-human gene transfer is possible. Lateral gene transfer can be a potential factor for cancer development.[28] Previously, it was thought that bacteria-to-human gene transfer was a rare event, but it is now recognized as a category of gene coding for approximately 223 genes that have been identified during the human genome project, which has similarity with bacteria, but no comparable similarity with worm, yeast, fly, or any other non-vertebrate eukaryotes. Therefore, it suggests lateral gene transfer of such genes. Obligate and invasive nature of *S. typhimurium* coupled with detection of strong inter-strain lateral gene transfer frequency suggests potential involvement of *S. typhimurium* lateral gene transfer in colon cancer etiology.[21,32] The chaperones contain cancer antagonist properties by working as genetic buffers, which stabilizes the usual phenotype.[27] Therefore, the alteration in chaperones can act as a potential factor for cancer development. Although this study provides a background for *S. typhimurium* protein targeting in the host cell, the validation of these results is extremely important before any conclusion can be drawn.

Conclusion

Although there is considerable, well-justified speculation that *S. typhimurium* infection might play an important role in colon cancer etiology, this study provides, to the best of our knowledge, the first demonstration of a nuclear targeting of *S. typhimurium* proteins in host cells, which may be significantly linked with colon cancer development. The mechanism of *S. typhimurium* in the development of cancer is still unclear, and the studies proposing and opposing this relation are enormous. The current in-silico work showed that various *S. typhimurium* proteins targeted to host cell nucleus which may have a profound impact on colon cancer etiology. These proteins can potentially affect the normal functioning of different pathways of the host cell. The current understanding of *S. typhimurium* infection and colon cancer is very limited, and this study makes a valuable contribution to help us understand the relation between *S. typhimurium* infection and colon cancer. Therefore, this study provides a detailed perspective that can help narrow down the targets for studying the important role of different proteins of *S. typhimurium* in development of colon cancer in the future. Our work may open new avenues of research and aid in the management of colon cancer.

Acknowledgments

Disclosure

The authors have declared that they have no conflicts of interest relative to the present study.

References

1. Apweiler R, Bairoch A, Wu CH, et al. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res*. 2004;32:D115–119. doi:10.1093/nar/gkh131
2. Arthur JC, Perez-Chanona E, Muhlbauer M, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science*. 2012;338:120–123. doi:10.1126/science.1224820
3. Berger H, Marques MS, Zietlow R, Meyer TF, Machado JC, Figueiredo C. Gastric cancer pathogenesis. *Helicobacter*. 2016;21 (Suppl 1):34–38. doi:10.1111/hel.12338
4. Boccellato F, Meyer TF. Bacteria moving into focus of human cancer. *Cell Host Microbe*. 2015;17(6):728–730. doi:10.1016/j.chom.2015. 05.016
5. Chan AT, Giovannucci EL. Primary prevention of colorectal cancer. *Gastroenterology*. 2010;138(6):2029–2043 e2010. doi:10.1053/j. gastro.2010.01.057
6. Chumduri C, Gurumurthy RK, Zietlow R, Meyer TF. Subversion of host genome integrity by bacterial pathogens. *Nat Rev Mol Cell Biol*. 2016;17(10):659–673. doi:10.1038/nrm.2016.100
7. Coghill AE, Hildesheim A. Epstein-Barr virus antibodies and the risk of associated malignancies: review of the literature. *Am J Epidemiol*. 2014;180:687–695. doi:10.1093/aje/kwu176
8. Cokol M, Nair R, Rost B. Finding nuclear localization signals. *EMBO Rep*. 2000;1(5):411–415. doi:10.1093/embo-reports/kvd092
9. Cunningham C, Nemunaitis J. A phase I trial of genetically modified Salmonella typhimurium expressing cytosine deaminase (TAPET-CD, VNP20029) administered by intratumoral injection in combination with 5-fluorocytosine for patients with advanced or metastatic cancer. Protocol no: CL-017. Version: april 9, 2001. *Hum Gene Ther*. 2001;12(12):1594–1596.
10. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell*. 1990;61(5):759–767. doi:10.1016/0092-8674(90)90186-I
11. Figueira R, Holden DW. Functions of the Salmonella pathogenicity island 2 (SPI-2) type III secretion system effectors. *Microbiology*. 2012;158:1147–1161. doi:10.1099/mic.0.058115-0
12. Gagnaire A, Nadel B, Raoult D, Neefjes J, Gorvel JP. Collateral damage: insights into bacterial mechanisms that predispose host cells to cancer. *Nat Rev Microbiol*. 2017;15(2):109–128. doi:10.1038/nrmicro.2016.171
13. Gallo A, Miele M, Badami E, Conaldi PG. Molecular and cellular interplay in virus-induced tumors in solid organ recipients. *Cell Immunol*. 2018.
14. Garg A, Bhasin M, Raghava GP. Support vector machine-based method for subcellular localization of human proteins using amino acid compositions, their order, and similarity search. *J Biol Chem*. 2005;280(15):14427–14432. doi:10.1074/jbc.M411789200
15. Gorlich D, Mattaj IW. Nucleocytoplasmic transport. *Science*. 1996;271(5255):1513–1518. doi:10.1126/science.271.5255.1513
16. Hsieh P, Yamane K. DNA mismatch repair: molecular mechanism, cancer, and ageing. *Mech Ageing Dev*. 2008;129(7–8):391–407. doi:10.1016/j.mad.2008.02.012
17. Huang SC, Huang SF, Chen YT, et al. Overexpression of MutL homolog 1 and MutS homolog 2 proteins have reversed prognostic implications for stage I-II colon cancer patients. *Biomed J*. 2017;40:39–48. doi:10.1016/j.bj.2017.01.004
18. Kahrstrom CT. Bacterial pathogenesis: E. coli claims the driving seat for cancer. *Nat Rev Cancer*. 2012;12:658–659. doi:10.1038/nrc3363
19. Karin M. Too many transcription factors: positive and negative interactions. *New Biol*. 1990;2:126–131.
20. Kattan Z, Marchal S, Brunner E, et al. Damaged DNA binding protein 2 plays a role in breast cancer cell growth. *PLoS One*. 2008;3:e2002. doi:10.1371/journal.pone.0002002
21. Kawaguchi K, Murakami T, Suetsugu A, et al. High-efficacy targeting of colon-cancer liver metastasis with Salmonella typhimurium A1-R via intra-portal-vein injection in orthotopic nude-mouse models. *Oncotarget*. 2017;8:19065–19073. doi:10.18632/oncotarget.12227
22. Khan S. Potential role of Escherichia coli DNA mismatch repair proteins in colon cancer. *Crit Rev Oncol Hematol*. 2015;96 (3):475–482. doi:10.1016/j.critrevonc.2015.05.002
23. Khan S, Imran A, Khan AA, Abul Kalam M, Alshamsan A. Systems biology approaches for the prediction of possible role of Chlamydia pneumoniae proteins in the etiology of lung cancer. *PLoS One*. 2016;11:e0148530. doi:10.1371/journal.pone.0148530
24. Khan S, Zakariah M, Palaniappan S. Computational prediction of Mycoplasma hominis proteins targeting in nucleus of host cell and their implication in prostate cancer etiology. *Tumour Biol*. 2016;37:10805–10813. doi:10.1007/s13277-016-4970-9
25. Kosugi S, Hasebe M, Entani T, Takayama S, Tomita M, Yanagawa H. Design of peptide inhibitors for the importin alpha/beta nuclear import pathway by activity-based profiling. *Chem Biol*. 2008;15 (9):940–949. doi:10.1016/j.chembiol.2008.07.019
26. Kosugi S, Hasebe M, Tomita M, Yanagawa H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci U S A*. 2009;106:10171–10176. doi:10.1073/pnas.0900604106

27. Kroll J. Molecular chaperones and the epigenetics of longevity and cancer resistance. *Ann N Y Acad Sci.* 2007;1100(1):75–83. doi:10.1196/annals.1395.006

28. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001;409:860–921.

29. Lange A, Mills RE, Lange CJ, Stewart M, Devine SE, Corbett AH. Classical nuclear localization signals: definition, function, and interaction with importin alpha. *J Biol Chem.* 2007;282:5101–5105. doi:10.1074/jbc.R600026200

30. Libermann TA, Zerbini LF. Targeting transcription factors for cancer gene therapy. *Curr Gene Ther.* 2006;6(1):17–33. doi:10.2174/156652306775515501

31. Liu X, Lu R, Xia Y, Wu S, Sun J. Eukaryotic signaling pathways targeted by Salmonella effector protein AvrA in intestinal infection in vivo. *BMC Microbiol.* 2010;10:326. doi:10.1186/1471-2180-10-326

32. Lu R, Bosland M, Xia Y, Zhang YG, Kato I, Sun J. Presence of Salmonella AvrA in colorectal tumor and its precursor lesions in mouse intestine and human specimens. *Oncotarget.* 2017;8 (33):55104–55115. doi:10.18632/oncotarget.19052

33. Lu R, Wu S, Zhang YG, et al. Enteric bacterial protein AvrA promotes colonic tumorigenesis and activates colonic beta-catenin signaling pathway. *Oncogenesis.* 2014;3(6):e105. doi:10.1038/oncsis.2014.20

34. Mager DL. Bacteria and cancer: cause, coincidence or cure? A review. *J Transl Med.* 2006;4:14. doi:10.1186/1479-5876-4-14

35. Maw MK, Fujimoto J, Tamaya T. Overexpression of inhibitor of DNA-binding (ID)-1 protein related to angiogenesis in tumor advancement of ovarian cancers. *BMC Cancer.* 2009;9:430. doi:10.1186/1471-2407-9-430

36. McClelland M, Sanderson KE, Spieth J, et al. Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. *Nature.* 2001;413:852–856. doi:10.1038/35101614

37. Mote J Jr., Reines D. Recognition of a human arrest site is conserved between RNA polymerase II and prokaryotic RNA polymerases. *J Biol Chem.* 1998;273:16843–16852. doi:10.1074/jbc.273.27.16843

38. Mughini-Gras L, Schaapveld M, Kramers J, et al. Increased colon cancer risk after severe Salmonella infection. *PLoS One.* 2018;13: e0189721. doi:10.1371/journal.pone.0189721

39. Nemunaitis J, Cunningham C, Senzer N, et al. Pilot trial of genetically modified, attenuated Salmonella expressing the E. coli cytosine deaminase gene in refractory cancer patients. *Cancer Gene Ther.* 2003;10:737–744.

40. Obmolova G, Ban C, Hsieh P, Yang W. Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature.* 2000;407:703–710. doi:10.1038/35037509

41. Parry L, Clarke AR. The roles of the Methyl-CpG binding proteins in cancer. *Genes Cancer.* 2011;2(6):618–630. doi:10.1177/1947601911418499

42. Pierleoni A, Martelli PL, Fariselli P, Casadio R. BaCelLo: a balanced subcellular localization predictor. *Bioinformatics.* 2006;22(14):e408–416. doi:10.1093/bioinformatics/btl222

43. Polk DB, Peek RM Jr. Helicobacter pylori: gastric cancer and beyond. *Nat Rev Cancer.* 2010;10:403–414. doi:10.1038/nrc2857

44. Pundir S, Martin MJ, O'Donovan C. UniProt protein knowledgebase. *Methods Mol Biol.* 2017;1558:41–55.

45. Rathi B, Sarangi AN, Trivedi N. Genome subtraction for novel target definition in Salmonella typhi. *Bioinformation.* 2009;4:143–150. doi:10.6026/97320630004143

46. Reinhardt A, Hubbard T. Using neural networks for prediction of the subcellular location of proteins. *Nucleic Acids Res.* 1998;26:2230–2236. doi:10.1093/nar/26.9.2230

47. Samaras V, Rafailidis PI, Mourtzoukou EG, Peppas G, Falagas ME. Chronic bacterial and parasitic infections and cancer: a review. *J Infect Dev Ctries.* 2010;4:267–281. doi:10.3855/jidc.819

48. Scanu T, Spaapen RM, Bakker JM, et al. Salmonella manipulation of host signaling pathways provokes cellular transformation associated with gallbladder carcinoma. *Cell Host Microbe.* 2015;17:763–774. doi:10.1016/j.chom.2015.05.002

49. Shen HB, Chou KC. Hum-mPLoc: an ensemble classifier for large-scale human protein subcellular location prediction by incorporating samples with multiple sites. *Biochem Biophys Res Commun.* 2007;355:1006–1011. doi:10.1016/j.bbrc.2007.02.071

50. Shen HB, Chou KC. A top-down approach to enhance the power of predicting human protein subcellular localization: hum-mPLoc 2.0. *Anal Biochem.* 2009;394:269–274. doi:10.1016/j.ab.2009.07.046

51. Su EC, Chang JM, Cheng CW, Sung TY, Hsu WL. Prediction of nuclear proteins using nuclear translocation signals proposed by probabilistic latent semantic indexing. *BMC Bioinformatics.* 2012;13 Suppl 17:S13. doi:10.1186/1471-2105-13-S17-S13

52. Sweetser D, Nonet M, Young RA. Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. *Proc Natl Acad Sci U S A.* 1987;84:1192–1196. doi:10.1073/pnas.84.5.1192

53. Toller IM, Neelsen KJ, Steger M, et al. Carcinogenic bacterial pathogen Helicobacter pylori triggers DNA double-strand breaks and a DNA damage response in its host cells. *Proc Natl Acad Sci U S A.* 2011;108:14944–14949. doi:10.1073/pnas.1100959108

54. Wang G, Gao J, Huang H, et al. Expression of a LINE-1 endonuclease variant in gastric cancer: its association with clinicopathological parameters. *BMC Cancer.* 2013;13:265. doi:10.1186/1471-2407-13-265

55. Wang J, Lu R, Fu X, et al. Novel regulatory roles of wnt1 in infection-associated colorectal cancer. *Neoplasia.* 2018;20 (5):499–509. doi:10.1016/j.neo.2018.03.001

56. Wei EK, Giovannucci E, Wu K, et al. Comparison of risk factors for colon and rectal cancer. *Int J Cancer.* 2004;108:433–442. doi:10.1002/ijc.11540

57. Wu X, Zhu Z, Li W, et al. Chromodomain helicase DNA binding protein 5 plays a tumor suppressor role in human breast cancer. *Breast Cancer Res.* 2012;14:R73. doi:10.1186/bcr3182

58. Zur Hausen H. The search for infectious causes of human cancers: where and why. *Virology.* 2009;392:1–10. doi:10.1016/j.virol.2009.06.001