

RESEARCH ARTICLE

# Viruses in case series of tumors: Consistent presence in different cancers in the same subject

Laila Sara Arroyo Mühr, Maria Hortlund, Zurab Bzhalava, Sara Nordqvist Kleppe, Davit Bzhalava, Emilie Hultin, Joakim Dillner\*

Division of Pathology, Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden

\* [joakim.dillner@ki.se](mailto:joakim.dillner@ki.se)



**OPEN ACCESS**

**Citation:** Arroyo Mühr LS, Hortlund M, Bzhalava Z, Nordqvist Kleppe S, Bzhalava D, Hultin E, et al. (2017) Viruses in case series of tumors: Consistent presence in different cancers in the same subject. PLoS ONE 12(3): e0172308. doi:10.1371/journal.pone.0172308

**Editor:** Andrzej T. Slominski, University of Alabama at Birmingham, UNITED STATES

**Received:** November 11, 2016

**Accepted:** February 2, 2017

**Published:** March 3, 2017

**Copyright:** © 2017 Arroyo Mühr et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data are available from the NCBI SRA database with the following accession numbers: SRR5209762 SRP098091; SRR5209761 SRP098091; SRR5209760 SRP098091; SRR5209759 SRP098091; SRR5209758 SRP098091; SRR5209757 SRP098091; SRR5209756 SRP098091; SRR5209755 SRP098091; SRR5209754 SRP098091; SRR5209753 SRP098091; SRR5209752 SRP098091; SRR5209751 SRP098091; SRR5209750 SRP098091; SRR5209749 SRP098091; SRR5209748

## Abstract

Studies investigating presence of viruses in cancer often analyze case series of cancers, resulting in detection of many viruses that are not etiologically linked to the tumors where they are found. The incidence of virus-associated cancers is greatly increased in immunocompromised individuals. Non-melanoma skin cancer (NMSC) is also greatly increased and a variety of viruses have been detected in NMSC. As immunosuppressed patients often develop multiple independent NMSCs, we reasoned that viruses consistently present in independent tumors might be more likely to be involved in tumorigenesis. We sequenced 8 different NMSCs from 1 patient in comparison to 8 different NMSCs from 8 different patients. Among the latter, 12 different virus sequences were detected, but none in more than 1 tumor each. In contrast, the patient with multiple NMSCs had human papillomavirus type 15 and type 38 present in 6 out of 8 NMSCs.

## Introduction

The International Agency for Research on Cancer has classified >16% of all cancers in humans as caused by infections [1]. The virus-associated cancers have a greatly increased incidence in immunosuppressed individuals with an impaired ability to control virus infections, such as AIDS patients or solid organ transplant recipients [2–6]. The cancer that has the most highly increased incidence among the immunosuppressed is non-melanoma skin carcinoma (NMSC) [7–9], but does not have an established infectious etiology. A number of infections have been found in NMSC, with Human papillomavirus (HPV) being the most commonly found [10–15], but it is unclear whether any one of the infections detected so far are etiologically linked to the tumor.

Sequencing of all nucleic acids present in a sample is today a routine procedure [16–19]. The human sequences can be removed in the bioinformatics analysis of the sequences, leaving the metagenome (non-human genomes of e.g. bacteria and viruses) for study [20–22]. Virus detection that is dependent on prior knowledge of sequences, for example PCR with specific primer sets, can give a biased representation of which viruses that are present [23] and a simple

SRP098091; SRR5209747 SRP098091;  
SRR5209746 SRP098091; SRR5209745  
SRP098091

**Funding:** This work was supported by the Swedish Research Council, grant number K2015-57X-21044-07-3 ([www.vr.se](http://www.vr.se)). The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

sequencing of all nucleic acids in a sample gives a more unbiased representation of the viruses that are present.

Sequencing of the viral metagenome in skin samples reported that >95% of the virus sequences found belonged to the Papillomaviridae family, mostly to the  $\beta$ - and  $\gamma$ -genera [24], although recent improvements in bioinformatics using hidden Markov models now allow detection of a broader array of viruses in such samples [25]. A multitude of HPV types is present in skin lesions as well as on healthy skin, both among immunocompetent and immunosuppressed patients [11, 12, 26–29]. The description of which viruses that is most frequently present appears to be highly dependent on the different PCR primers used for HPV detection [12–15]. Sequencing of specimens without prior PCR is a more unbiased approach to detect viruses, even those which are not amplifiable by known PCR systems (2).

There are no fully adequate controls to use for case series of cancer tissue and it is therefore difficult to even guess if detection of a virus in a tumor suggests an etiological involvement or not. We reasoned that viruses consistently present in a series of independent tumors from the same patient might be more likely to be involved in tumorigenesis than viruses only found non-systematically in only some specimens. We therefore set out to test if there were indeed any virus types that were consistently found in such tumor series. Similar reasoning has been discussed before at a time point where HPV type-specific PCR-based approaches were used to detect HPV, however no single HPV type seemed to be consistently present in skin lesions of transplant recipients using biased PCR-based detection [30].

## Materials and methods

### Patients

We linked the 13429 patients having received a solid organ transplant in Sweden with the Swedish Cancer Registry and found 11127 tumors diagnosed >6 months after the transplantation. For this study, we selected 8 available diagnostic specimens from 8 different NMSCs (squamous cell carcinomas) that all had occurred in the same patient (diagnosed over a 10-year period, from 1993 to 2003). The localizations of the tumors were from “other and unspecified parts of the face” (6 specimens), “scalp and neck” (1 specimen) and “upper limb” (1 specimen). As controls, we identified 8 diagnostic blocks from NMSC (squamous cell carcinomas) that had occurred in 8 different transplant recipients. The matching of “controls” to the case was performed according to 1) gender (male), 2) diagnosis (ICD7 code 191 “other malignant neoplasm of skin” including the forth digit “localization of the body”), 3) county of residence, 4) year of diagnosis (+/-10 years) and 5) age of the patient at diagnosis (+/-10 years).

The study was approved by the Ethical Review Board of the Stockholm region (Sweden) and has been conducted according to the principles expressed in the Declaration of Helsinki.

### Sample preparation

The FFPE blocks were sectioned at a quality-assured, certified laboratory (HistoCentre, Gothenburg, Sweden) where the process has been repeatedly validated to ensure uniform and contamination-free sectioning from blocks. Empty paraffin blocks were also sectioned in-between each sample to be used as negative controls.

Four 5 $\mu$ m sections of each formalin fixed paraffin embedded specimen (both tumor and blank paraffin blocks) were subjected to serial DNA extraction using the ExScale Automated DNA/RNA purification kit (art. no. ES-K110210FP) on a Magtration® 12GC UV at ExScale Biospecimen Solutions AB (Uppsala, Sweden), following the manufacturer’s instructions.

Analysis of the presence of the *hemoglobin subunit beta* gene with real-time PCR was performed to assess sample adequacy [31]. *Hemoglobin subunit beta* negative tumor specimens

were re-analyzed to confirm negativity. If negativity was confirmed, specimens were excluded for further analysis.

All individual tumor blocks positive for *Hemoglobin subunit beta* together with 2 negative controls (1 water control and 1 blank paraffin control) were amplified with whole genome amplification (WGA) using the Illustra™ Ready-To-Go™ GenomiPhi™ DNA Amplification Kit (GE Health Care, United Kingdom). The manufacturer's instructions were followed, with these modifications: 2.5 µL of sample were diluted with 22.5 µL PCR-Grade water and 25 µL denaturation buffer. After incubation at 95°C for 3 min and cooling on ice, 50 µL of the samples were added to the Ready-To-Go GenomiPhi lyophilized cake. After a further incubation at 30°C for 7 h and inactivation at 65°C for 10 min, products were diluted in the ratio 1:2 in PCR-Grade water and quantified with QuantiFluor-ST (Promega, United States).

## Illumina sequencing

50 ng of genomic DNA per sample was tagged with 2 unique indices, as described (Nextera DNA Sample Preparation kit, user guide revision B (Illumina)). The individual libraries were validated, normalized to 4 nM, pooled, denatured and diluted, resulting in a 2.6 pM DNA solution, which was spiked at 1% PhiX control and the library pool was sequenced paired-end 151 +151 cycles using NextSeq 500 High Output reagent kit (Illumina) as described in the user guides "Denaturing and Diluting Libraries for the NextSeq 500" revision A and "NextSeq 500 kit Reference Guide" revision F.

## Bioinformatics

The index tags were used to assign the sequence reads to the individual samples. Sequences were trimmed according to their Phred quality scores [32] and then screened against the human reference genome hg19, as well as bacterial, phage and vector sequences (obtained from GenBank using BWA-MEM [33]) and reads with >95% identity over 75% of their length to human DNA were removed. The remaining sequences were normalized (<http://ged.msu.edu/papers/2012-diginorm>) and then assembled using the Trinity [34], SOAPdenovo-Trans (<http://soap.genomics.org.cn/>) and IDBA-UD [35] assemblers into contiguous sequences (contigs). Reads before assembly were re-mapped to assembled contigs. We used several assembly algorithms and re-mapping of all singleton reads to assembled contigs in order to validate assembly results [20, 36].

Taxonomic classification of assembled contigs used GenBank and paracel blast ([www.paracel.com](http://www.paracel.com)) to classify into 1) previously known sequences, 2) related to previously known sequences, or 3) unrelated to any previously known sequences. Papillomavirus-related contigs and singletons were checked to identify possible artifactual "chimeric" sequences (contigs containing sequences originating from different viruses), as described [20]. The sequence that aligned to its most closely related sequence in GenBank was divided into three equal segments. If at least one segment differed in similarity to the corresponding overlapping parts with more than 5% the sequence was considered a "possible chimera".

Reads classified as "unrelated to any previously known sequences" were also analyzed with HMMER (the hmmsearch algorithm) [37] to search for distant homologs of already known viruses among the contigs. HMMER is designed to search sequence databases and identify remote homologues by implementing profile hidden Markov models. As reference database, we used the database constructed by Peter Skewes-Cox et al. (<http://derisilab.ucsf.edu/software/vFam>) [38], which includes viral profile hidden Markov models ("vFams") from all the virally annotated proteins in RefSeq [38]. The HMMER algorithm calculates the E-value to

**Table 1. Taxonomic assignment of sequencing reads.**

	8 NMSCs <sup>a</sup> same patient		8 NMSCs <sup>a</sup> different patients		All	
	Reads	%	Reads	%	Reads	%
Bacteria	71196424	16.01	31188714	7.94	122696992	12.65
Human	353941360	79.6	342072130	87.13	754180186	77.74
Other	16171088	3.64	15083268	3.84	53159844	5.48
Unknown	3316424	0.75	4245870	1.08	40072660	4.13
Virus	38416	0.01	38	0.00	45970	0.00
TOTAL	444663712	100	392590020	100	970155652	100

The number of reads detected by taxonomic division in the cases of NMSCs from the same patient and in the 8 NMSCs from different patients, as well as the total number of reads in the sequencing run. The classification “Other” included reads that mapped to contigs with similarities to other sequences present in GenBank (e.g. plants, synthetic constructs, polymerase background reactivity). Reads mapped to contigs that did not give any hit to genomes deposited in GenBank were classified as “Unknown”. Reads detected in negative controls (water and blank paraffin block) as well as undetermined reads (reads failing the index identification) are only included in the “All” column.

<sup>a</sup>NMSCs: Non-melanoma skin carcinoma.

doi:10.1371/journal.pone.0172308.t001

order top sequences. We selected HHMER hits with e-value less than 1e-5. If sequences had hits with several different virus families, the hit with the lowest E-value was used.

Analyses used in-house R ([www.R-project.org/](http://www.R-project.org/)), python ([www.python.org](http://www.python.org)) and bash (<http://www.gnu.org/software/bash/>) scripts running on a high performance (40 core, 2 TB RAM) Linux server. Specimens were classified as positive for viruses if at least 5 reads were detected.

## Results

Eight different NMSC from the case and 8 NMSC from the different patients were sequenced using the Illumina technology. There were ~970 million sequencing reads, from which <0.01% belonged to viral sequences (Table 1).

The Blast-based bioinformatics pipeline revealed 45970 viral reads, of which 38416 (83.6% of viral reads) were detected in the 8 different NMSC tumors belonging to the case and only 38 reads were detected in the 8 NMSC specimens belonging to the different patients (Table 2). Most viral reads (26308 viral reads) were from Human Herpesvirus 7 (HHV-7), which was found in only one NMSC specimen. Human papillomavirus (HPV) was detected in 6/8 NMSC specimens retrieved from the case patient (all 6 of them had a double infection with both HPV 15 and HPV 38). Only one NMSC among the controls had HPV detected (a single HPV 38 infection). The same “Faustovirus-related” apparently environmental virus was found in both case specimens, control specimen as well as in the water control with 2348, 20 and 5562 reads, respectively (Table 2).

All reads that mapped to contigs that did not give any hit to genomes deposited in GenBank were classified as “unknown”, representing 4.13% of total reads (Table 1). Searching for distant homologs of already known viruses among these contigs using profile hidden Markov models revealed 23 (previously not described) viral contigs (designated as SE-sequences) related to up to 5 viral clusters (Table 2). None of the 23 SE-sequences were detected in more than 2 specimens each and only 2/5 clusters (Baculoviridae and Marseilleviridae) were not detected in the negative controls. The DNA viral contigs detected with the HHMER analysis (designated SE881-SE902, SE871) have been deposited in GenBank (accession numbers KY006985-KY007007) and the deep sequencing data is deposited at NIH Sequence Read Archive (BioSample accessions SAMN06275823-SAMN06275840).

**Table 2. Classification of viral reads.**

Virus family	8 NMSCs <sup>a</sup> same patient		8 NMSCs <sup>a</sup> different patients		Water	Paraffin blank block	All
	Reads	Pos <sup>b</sup> Specimens	Reads	Pos <sup>b</sup> Specimens	Reads	Reads	Reads
Papillomaviridae <sup>B</sup>	9760	6/8	18	1/8	0	0	10096
HPV-15	8424	6/8	4	0/8	0	0	8698
HPV-38	1336	6/8	14	1/8	0	0	1398
Herpesviridae <sup>B</sup>	26308	1/8	0	0/8	0	0	27584
HHV-7 <sup>B</sup>	26308	1/8	0	0/8	0	0	27584
Related to Faustovirus <sup>B</sup>	2348	2/8	20	1/8	5652	0	8290
Baculoviridae <sup>H</sup>	4258	2/8	6134	1/8	0	0	10790
SE-882, 884, 890, 898	0	0/8	6134	1/8	0	0	6356
SE-896	3568	1/8	0	0/8	0	0	3722
SE-899	690	2/8	0	0/8	0	0	712
Marseilleviridae <sup>H</sup>	0	0/8	2210	1/8	0	0	2292
SE-895	0	0/8	2210	1/8	0	0	2292
Mimiviridae <sup>H</sup>	618	1/8	882	2/8	254	758	2606
SE-883, 893, 897	6	0/8	880	2/8	152	66	1152
SE-894	612	1/8	0	0/8	0	0	632
SE-885, 886, 888, 900, 902	0	0/8	2	0/8	102	692	822
Phycodnaviridae <sup>H</sup>	0	0/8	26	1/8	0	832	884
SE-881	0	0/8	0	0/8	0	832	854
SE-889	0	0/8	26	1/8	0	0	30
Unassigned <sup>H</sup>	8950	3/8	1056	1/8	36	0	10376
SE-887, 901	0	0/8	1056	1/8	0	0	1094
SE-891	32	1/8	0	0/8	0	0	32
SE-871	8918	2/8	0	0/8	0	0	9214
SE-892	0	0/8	0	0/8	36	0	36
Virus <sup>B</sup>	38416	7/8	38	1/8	5652	0	45970
Virus <sup>H</sup>	13826	5/8	10308	5/8	290	1590	26948
Total virus	52242	8/8	10346	6/8	5942	1590	72918

Description of viral sequences detected using paracel blast (<sup>B</sup>) and HMMER (<sup>H</sup>) bioinformatics algorithms. Unassigned<sup>H</sup> classification corresponds to clusters which are constructed from viral sequences that do not share one viral family or do not belong to any viral family. Undetermined reads (reads failing the index identification) are only included in the “All” column.

<sup>a</sup>NMSC: Non-melanoma skin carcinoma

<sup>b</sup>Pos: Positive.

doi:10.1371/journal.pone.0172308.t002

## Discussion

We report that sequencing of tumor tissues from different cancers occurring up to 10 years apart in the same individual, to a large extent tended to detect the same viruses present in multiple different tumors. While this finding does in itself not imply causality, it does suggest that viruses that are systematically detected in multiple tumors from the same patient might be more likely to be involved in tumorigenesis than viruses only found non-systematically in only some specimens.

The study is based on only a few patients and more patients would need to be sequenced to obtain more large-scale data. Although deep sequencing is today a routine, it is still costly and more large-scale studies would probably need to await less expensive technologies to be enabled. Another limitation is that the study compared only different cases of SCC, whereas

benign tissues were not analyzed. As the study design was based on specimens from pathology archives, only tumors (and not benign tissues) were available.

The majority of viral reads detected (37.8% of total viral reads), were from HHV-7, a beta-herpesvirus known to infect almost all children by the age of three years with subsequent life-long persistence, but all the HHV-7 reads were from a single specimen. The use of sequencing for virus detection is not dependent on prior knowledge of sequences, and therefore a more unbiased representation of viruses will be detected. However, the fact that we before sequencing used whole genome amplification (WGA) to increase the amount of DNA in the sample means that the number of reads obtained is not directly correlated to the amount of viral populations present [39]. This phenomenon is well described and is probably attributable to preferential amplification of genomic regions during initial WGA priming events [16, 40, 41]. The significance of the large amount of HHV-7 reads in a single specimen is therefore unclear.

Considering presence in multiple tumors, HPV was the most common virus found. Double infection of HPV 15 and 38 was detected in 6/8 NMSC belonging to the same patient, whereas the NMSCs from 8 different patients were HPV-positive only in 1/8 tumors (a single infection with HPV38). Both HPV 15 and 38 belong to the genus Beta-papillomavirus and are commonly found in patients with epidermodysplasia verruciformis. Further studies are needed to resolve the biological significance of the finding of HPV15 and 38 in multiple tumors from the same patient, but it can be concluded that as these cancers occurred up to 10 years apart the double infection with HPV15 and 38 must have persisted for many years in this subject. Persistence of HPV over time is known to be of importance for tumor maintenance and progression.

A large part of the sequencing reads from *de novo* sequencing projects are classified as “Unknown” due to incompleteness of public sequence databases or drawbacks of next generation sequencing technologies such as short read lengths and sequencing errors [20, 21]. The Blast-based bioinformatics pipeline revealed ~45000 viral reads, while the number of “Unknown” reads reached ~4M. Therefore, the more sensitive HHMER algorithms were used in order to search for distant homologs of already known viruses among the contigs classified as “Unknown” [25]. HHMER-based analysis identified 23 previously unknown viral contigs with up to 26948 viral reads. Each contig within a cluster was assigned a different SE-number (e.g: SE898) if its sequence did not overlap with other contigs related to the same cluster. Although the different SE-sequences were not overlapping, it is still possible that different contigs from the same cluster may belong to the same virus.

Viral sequences belonging to Mimiviridae, Phycodnaviridae, “related to Faustovirus”, and “Unassigned” (distantly related to viruses previously described, but not given any taxonomic designation yet), were detected in not more than 2 specimens each as well as in the water and/or the blank paraffin block used as negative controls. These sequences might therefore represent environmental deposit, highlighting the importance of including negative controls whenever performing metagenomic sequencing.

## Conclusions

In this study, deep sequencing was used for unbiased virus detection in NMSC specimens. Consecutive specimens from the same patient over several years and from different locations identified presence of HPV (types 15 and 38) in most of the specimens (6 out of 8), whereas only 1 out of 8 specimens from different patients was HPV-positive (type 38).

## Author Contributions

**Conceptualization:** JD EH.

**Data curation:** LSAM MH SNK ZB DB EH.

**Formal analysis:** ZB DB EH LSAM.

**Funding acquisition:** JD.

**Investigation:** LSAM MH ZB DB EH.

**Methodology:** JD EH ZB DB.

**Project administration:** JD EH LSAM.

**Resources:** MH SNK ZB DB EH JD.

**Software:** ZB DB.

**Supervision:** JD EH.

**Validation:** EH.

**Visualization:** LSAM DB EH.

**Writing – original draft:** LS AM.

**Writing – review & editing:** LSAM MH SNK ZB DB EH JD.

## References

1. de Martel C, Ferlay J, Franceschi S, Vignat J, Bray F, Forman D, et al. Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *The Lancet Oncology*. 2012 Jun; 13(6):607–15. doi: [10.1016/S1470-2045\(12\)70137-7](https://doi.org/10.1016/S1470-2045(12)70137-7) PMID: [22575588](https://pubmed.ncbi.nlm.nih.gov/22575588/)
2. Grulich AE, van Leeuwen MT, Falster M, Vajdic CM. Incidence of cancers in people with HIV/AIDS compared with immunosuppressed transplant recipients: a meta-analysis. *Lancet*. 2007 Jul 7; 370(9581):59–67. doi: [10.1016/S0140-6736\(07\)61050-2](https://doi.org/10.1016/S0140-6736(07)61050-2) PMID: [17617273](https://pubmed.ncbi.nlm.nih.gov/17617273/)
3. Bhatia K, Shiels MS, Berg A, Engels EA. Sarcomas other than Kaposi sarcoma occurring in immunodeficiency: interpretations from a systematic literature review. *Curr Opin Oncol*. 2012 Sep; 24(5):537–46. doi: [10.1097/CCO.0b013e328355e115](https://doi.org/10.1097/CCO.0b013e328355e115) PMID: [22729152](https://pubmed.ncbi.nlm.nih.gov/22729152/)
4. Collett D, Mumford L, Banner NR, Neuberger J, Watson C. Comparison of the Incidence of Malignancy in Recipients of Different Types of Organ: A UK Registry Audit. *Am J Transplant*. 2010 Aug; 10(8):1889–96. doi: [10.1111/j.1600-6143.2010.03181.x](https://doi.org/10.1111/j.1600-6143.2010.03181.x) PMID: [20659094](https://pubmed.ncbi.nlm.nih.gov/20659094/)
5. van Leeuwen MT, Webster AC, McCredie MRE, Stewart JH, McDonald SP, Amin J, et al. Effect of reduced immunosuppression after kidney transplant failure on risk of cancer: population based retrospective cohort study. *Brit Med J*. 2010 Feb 11; 340.
6. Engels EA, Pfeiffer RM, Fraumeni JF, Kasiske BL, Israni AK, Snyder JJ, et al. Spectrum of Cancer Risk Among US Solid Organ Transplant Recipients. *Jama-J Am Med Assoc*. 2011 Nov 2; 306(17):1891–901.
7. Boukamp P. Non-melanoma skin cancer: what drives tumor development and progression? *Carcinogenesis*. 2005 Oct; 26(10):1657–67. doi: [10.1093/carcin/bgi123](https://doi.org/10.1093/carcin/bgi123) PMID: [15905207](https://pubmed.ncbi.nlm.nih.gov/15905207/)
8. Lindelof B, Sigurgeirsson B, Gabel H, Stern RS. Incidence of skin cancer in 5356 patients following organ transplantation. *Br J Dermatol*. 2000 Sep; 143(3):513–9. PMID: [10971322](https://pubmed.ncbi.nlm.nih.gov/10971322/)
9. Berg D, Otley CC. Skin cancer in organ transplant recipients: Epidemiology, pathogenesis, and management. *Journal of the American Academy of Dermatology*. 2002 Jul; 47(1):1–17; quiz 8–20. PMID: [12077575](https://pubmed.ncbi.nlm.nih.gov/12077575/)
10. Forslund O. Genetic diversity of cutaneous human papillomaviruses. *The Journal of general virology*. 2007 Oct; 88(Pt 10):2662–9. doi: [10.1099/vir.0.82911-0](https://doi.org/10.1099/vir.0.82911-0) PMID: [17872517](https://pubmed.ncbi.nlm.nih.gov/17872517/)
11. Asgari MM, Kiviat NB, Critchlow CW, Stern JE, Argenyi ZB, Raugi GJ, et al. Detection of human papillomavirus DNA in cutaneous squamous cell carcinoma among immunocompetent individuals. *The Journal of investigative dermatology*. 2008 Jun; 128(6):1409–17. doi: [10.1038/sj.jid.5701227](https://doi.org/10.1038/sj.jid.5701227) PMID: [18185530](https://pubmed.ncbi.nlm.nih.gov/18185530/)
12. Weissenborn SJ, Nindl I, Purdie K, Harwood C, Proby C, Breuer J, et al. Human papillomavirus-DNA loads in actinic keratoses exceed those in non-melanoma skin cancers. *The Journal of investigative dermatology*. 2005 Jul; 125(1):93–7. doi: [10.1111/j.0022-202X.2005.23733.x](https://doi.org/10.1111/j.0022-202X.2005.23733.x) PMID: [15982308](https://pubmed.ncbi.nlm.nih.gov/15982308/)

13. Vasiljevic N, Hazard K, Eliasson L, Ly H, Hunziker A, de Villiers EM, et al. Characterization of two novel cutaneous human papillomaviruses, HPV93 and HPV96. *The Journal of general virology*. 2007 May; 88 (Pt 5):1479–83. doi: [10.1099/vir.0.82679-0](https://doi.org/10.1099/vir.0.82679-0) PMID: [17412976](https://pubmed.ncbi.nlm.nih.gov/17412976/)
14. Hazard K, Eliasson L, Dillner J, Forslund O. Subtype HPV38b[FA125] demonstrates heterogeneity of human papillomavirus type 38. *International journal of cancer*. 2006 Sep 1; 119(5):1073–7. doi: [10.1002/ijc.21920](https://doi.org/10.1002/ijc.21920) PMID: [16557595](https://pubmed.ncbi.nlm.nih.gov/16557595/)
15. Vasiljevic N, Hazard K, Dillner J, Forslund O. Four novel human betapapillomaviruses of species 2 preferentially found in actinic keratosis. *The Journal of general virology*. 2008 Oct; 89(Pt 10):2467–74. doi: [10.1099/vir.0.2008/001925-0](https://doi.org/10.1099/vir.0.2008/001925-0) PMID: [18796715](https://pubmed.ncbi.nlm.nih.gov/18796715/)
16. Arroyo Muhr LS, Hultin E, Bzhalava D, Eklund C, Lagheden C, Ekstrom J, et al. Human papillomavirus type 197 is commonly present in skin tumors. *International journal of cancer*. 2015 Jun 1; 136 (11):2546–55. doi: [10.1002/ijc.29325](https://doi.org/10.1002/ijc.29325) PMID: [25388227](https://pubmed.ncbi.nlm.nih.gov/25388227/)
17. Johansson H, Bzhalava D, Ekstrom J, Hultin E, Dillner J, Forslund O. Metagenomic sequencing of "HPV-negative" condylomas detects novel putative HPV types. *Virology*. 2013 May 25; 440(1):1–7. doi: [10.1016/j.virol.2013.01.023](https://doi.org/10.1016/j.virol.2013.01.023) PMID: [23522725](https://pubmed.ncbi.nlm.nih.gov/23522725/)
18. Smelov V, Bzhalava D, Arroyo Muhr LS, Eklund C, Komyakov B, Gorelov A, et al. Detection of DNA viruses in prostate cancer. *Scientific reports*. 2016; 6:25235. doi: [10.1038/srep25235](https://doi.org/10.1038/srep25235) PMID: [27121729](https://pubmed.ncbi.nlm.nih.gov/27121729/)
19. Bzhalava D, Hultin E, Arroyo Muhr LS, Ekstrom J, Lehtinen M, de Villiers EM, et al. Viremia during pregnancy and risk of childhood leukemia and lymphomas in the offspring: Nested case-control study. *International journal of cancer*. 2016 May 1; 138(9):2212–20. doi: [10.1002/ijc.29666](https://doi.org/10.1002/ijc.29666) PMID: [26132655](https://pubmed.ncbi.nlm.nih.gov/26132655/)
20. Bzhalava D, Johansson H, Ekstrom J, Faust H, Moller B, Eklund C, et al. Unbiased Approach for Virus Detection in Skin Lesions. *Plos One*. 2013 Jun 28; 8(6).
21. Bzhalava D, Ekstrom J, Lysholm F, Hultin E, Faust H, Persson B, et al. Phylogenetically diverse TT virus viremia among pregnant women. *Virology*. 2012 Oct 25; 432(2):427–34. doi: [10.1016/j.virol.2012.06.022](https://doi.org/10.1016/j.virol.2012.06.022) PMID: [22819835](https://pubmed.ncbi.nlm.nih.gov/22819835/)
22. Ekstrom J, Muhr LS, Bzhalava D, Soderlund-Strand A, Hultin E, Nordin P, et al. Diversity of human papillomaviruses in skin lesions. *Virology*. 2013 Dec; 447(1–2):300–11. doi: [10.1016/j.virol.2013.09.010](https://doi.org/10.1016/j.virol.2013.09.010) PMID: [24210127](https://pubmed.ncbi.nlm.nih.gov/24210127/)
23. Arroyo Muhr LS, Bzhalava D, Lagheden C, Eklund C, Johansson H, Forslund O, et al. Does human papillomavirus-negative condylomata exist? *Virology*. 2015 Nov; 485:283–8. doi: [10.1016/j.virol.2015.07.023](https://doi.org/10.1016/j.virol.2015.07.023) PMID: [26318260](https://pubmed.ncbi.nlm.nih.gov/26318260/)
24. Foulongne V, Sauvage V, Hebert C, Dereure O, Cheval J, Gouilh MA, et al. Human skin microbiota: high diversity of DNA viruses identified on the human skin by high throughput sequencing. *Plos One*. 2012; 7(6):e38499. doi: [10.1371/journal.pone.0038499](https://doi.org/10.1371/journal.pone.0038499) PMID: [22723863](https://pubmed.ncbi.nlm.nih.gov/22723863/)
25. Bzhalava Z, Hultin E, Dillner J, Bzhalava D. Extended detection of viruses in humans using viral profile hidden Markov models. *Scientific reports*. 2016.
26. Forslund O, Lindelof B, Hradil E, Nordin P, Stenquist B, Kirnbauer R, et al. High prevalence of cutaneous human papillomavirus DNA on the top of skin tumors but not in "Stripped" biopsies from the same tumors. *The Journal of investigative dermatology*. 2004 Aug; 123(2):388–94. doi: [10.1111/j.0022-202X.2004.23205.x](https://doi.org/10.1111/j.0022-202X.2004.23205.x) PMID: [15245440](https://pubmed.ncbi.nlm.nih.gov/15245440/)
27. Forslund O, DeAngelis PM, Beigi M, Schjolberg AR, Clausen OP. Identification of human papillomavirus in keratoacanthomas. *Journal of cutaneous pathology*. 2003 Aug; 30(7):423–9. PMID: [12859739](https://pubmed.ncbi.nlm.nih.gov/12859739/)
28. de Koning MN, Struijk L, Bavinck JN, Kleter B, ter Schegget J, Quint WG, et al. Betapapillomaviruses frequently persist in the skin of healthy individuals. *The Journal of general virology*. 2007 May; 88(Pt 5):1489–95. doi: [10.1099/vir.0.82732-0](https://doi.org/10.1099/vir.0.82732-0) PMID: [17412978](https://pubmed.ncbi.nlm.nih.gov/17412978/)
29. Mackintosh LJ, de Koning MN, Quint WG, Ter Schegget J, Morgan IM, Herd RM, et al. Presence of beta human papillomaviruses in nonmelanoma skin cancer from organ transplant recipients and immunocompetent patients in the West of Scotland. *The British journal of dermatology*. 2009 Jul; 161(1):56–62. doi: [10.1111/j.1365-2133.2009.09146.x](https://doi.org/10.1111/j.1365-2133.2009.09146.x) PMID: [19416244](https://pubmed.ncbi.nlm.nih.gov/19416244/)
30. World Health Organization and International Agency for Research on Cancer. Human Papillomaviruses. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. vol. 90. Lyon: WHO press; 2007. p. 348.
31. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, et al. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. 1985. *Biotechnology*. 1992; 24:476–80. PMID: [1422056](https://pubmed.ncbi.nlm.nih.gov/1422056/)
32. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res*. 1998 Mar; 8(3):186–94.
33. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010 Mar 1; 26(5):589–95. doi: [10.1093/bioinformatics/btp698](https://doi.org/10.1093/bioinformatics/btp698) PMID: [20080505](https://pubmed.ncbi.nlm.nih.gov/20080505/)



34. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 2013 Aug; 8(8):1494–512. doi: [10.1038/nprot.2013.084](https://doi.org/10.1038/nprot.2013.084) PMID: [23845962](https://pubmed.ncbi.nlm.nih.gov/23845962/)
35. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics.* 2012 Jun 1; 28(11):1420–8. doi: [10.1093/bioinformatics/bts174](https://doi.org/10.1093/bioinformatics/bts174) PMID: [22495754](https://pubmed.ncbi.nlm.nih.gov/22495754/)
36. Meiring TL, Salimo AT, Coetzee B, Maree HJ, Moodley J, Hitzeroth II, et al. Next-generation sequencing of cervical DNA detects human papillomavirus types not detected by commercial kits. *Virology.* 2012 Aug 16;9.
37. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic acids research.* 2013 Jul; 41(12):e121. doi: [10.1093/nar/gkt263](https://doi.org/10.1093/nar/gkt263) PMID: [23598997](https://pubmed.ncbi.nlm.nih.gov/23598997/)
38. Skewes-Cox P, Sharpton TJ, Pollard KS, DeRisi JL. Profile hidden Markov models for the detection of viruses within metagenomic sequence data. *Plos One.* 2014; 9(8):e105067. doi: [10.1371/journal.pone.0105067](https://doi.org/10.1371/journal.pone.0105067) PMID: [25140992](https://pubmed.ncbi.nlm.nih.gov/25140992/)
39. Yilmaz S, Allgaier M, Hugenholtz P. Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nature methods.* 2010 Dec; 7(12):943–4. doi: [10.1038/nmeth1210-943](https://doi.org/10.1038/nmeth1210-943) PMID: [21116242](https://pubmed.ncbi.nlm.nih.gov/21116242/)
40. Dichosa AE, Fitzsimons MS, Lo CC, Weston LL, Preteska LG, Snook JP, et al. Artificial polyploidy improves bacterial single cell genome recovery. *Plos One.* 2012; 7(5):e37387. doi: [10.1371/journal.pone.0037387](https://doi.org/10.1371/journal.pone.0037387) PMID: [22666352](https://pubmed.ncbi.nlm.nih.gov/22666352/)
41. Wang J, Van Nostrand JD, Wu L, He Z, Li G, Zhou J. Microarray-based evaluation of whole-community genome DNA amplification methods. *Applied and environmental microbiology.* 2011 Jun; 77(12):4241–5. doi: [10.1128/AEM.01834-10](https://doi.org/10.1128/AEM.01834-10) PMID: [21498751](https://pubmed.ncbi.nlm.nih.gov/21498751/)