Next-generation sequencing fails to identify human virus sequences in cutaneous squamous cell carcinoma

Tina Ganzenmueller1, Yuri Yakushko1*, Jeanette Kluba1, Cornelia Henke-Gendo1,2, Ralf Gutzmer3 and Thomas F. Schulz1

1 Institute of Virology, Hannover Medical School, Hannover, Germany
2 Integrated Research and Treatment Center Transplantation (IFB-Tx), Hannover Medical School, Hannover, Germany
3 Skin Cancer Center Hanover, Department of Dermatology and Allergy, Hannover Medical School, Hannover, Germany

Nonmelanoma skin cancer (NMSC) shows a strongly increased incidence in solid organ transplant recipients (OTRs) and AIDS patients, suggesting an infectious etiology. The role of certain viruses, i.e., cutaneous human papillomaviruses (HPVs), in NMSC in immunosuppressed patients remains controversial. Merkel cell polyomavirus (MCPyV), which was recently identified using high-throughput sequencing, has been linked to cutaneous proliferations. Here, we aimed to identify novel or known viral sequences at the transcript level in cutaneous squamous cell carcinomas (SCCs) from OTR by using 454 high-throughput pyrosequencing, which can produce long reads (~400 bp) and thus is better suited for the analysis of unknown sequences than other sequencing platforms. cDNA libraries from three OTR SCC biopsies were generated and submitted to next-generation sequencing using a 454 platform. Bioinformatic analysis included digital transcriptome subtraction and—in parallel—reference mapping as an alternative way for depleting human sequences. All control sequences introduced for bioinformatics analysis were recovered correctly. Among 717,029 454-sequenced transcripts, nearly all identified viral reads were derived from phages. Bacterial sequences originated from the skin flora or environmental sources. Our study did not reveal any transcripts of known oncogenic or related unknown human viruses. These findings suggest that there is no abundant expression of known human viruses, or viruses with a high degree of homology to known human viruses, in cutaneous SCCs of OTR. Further studies are required to exclude the presence of viruses in NMSC, which cannot easily be identified on the basis of sequence homology to known viruses.

Key words: nonmelanoma skin cancer, next-generation sequencing, virus discovery, immunosuppression

Abbreviations: DTS: digital transcriptome subtraction; HPV: human papillomavirus; HPyV: human polyomavirus; MCPyV: Merkel cell polyomavirus; NMSC: nonmelanoma skin cancer; OTR: organ transplant recipient; SCC: squamous cell carcinoma

Additional Supporting Information may be found in the online version of this article

*T.G. and Y.Y. contributed equally to this work

Grant sponsor: Deutsche Forschungsgemeinschaft (DFG) Collaborative Research Centre; Grant numbers: SFB 900, SFB 566;

Grant sponsor: German Federal Ministry of Education and Research (BMBF) [Integrated Research and Treatment Center Transplantation (IFB-Tx)]

DOI: 10.1002/ijc.27581

History: Received 13 Nov 2011; Accepted 27 Mar 2012; Online 5 Apr 2012

Cornelia Henke-Gendo’s present address is: Institute for Medical Microbiology and Hospital Epidemiology, Hannover Medical School, Hannover, Germany

Correspondence to: Tina Ganzenmueller, Institute of Virology, Hannover Medical School, Carl-Neuberg-Straße 1, 30625 Hannover, Germany, Tel.: +49-511-532-4310, Fax: +49-511-532-8736, E-mail: ganzenmueller.tina@mh-hannover.de

Nonmelanoma skin cancer (NMSC) including cutaneous squamous cell carcinoma (SCC) is the most common skin cancer among solid organ transplant recipients (OTRs) and a significant cause of morbidity and mortality in these patients.1,2 The NMSC rates are markedly increased in OTR and AIDS patients compared to the general population.3,4 Furthermore, OTRs with NMSC tend to have multiple tumor lesions, a higher recurrence rate and the tumors are more likely to metastasize.5–7

Established risk factors for NMSC include exposure to ultraviolet radiation, fair color of the skin and immunosuppressive treatment.1,2 The question whether the increased NMSC rate in immunocompromised individuals—in particular OTR—could reflect a viral involvement in skin carcinogenesis has been investigated intensively. Although there is evidence that human papillomaviruses (HPVs), especially from species beta and gamma, could play a role in the development of NMSC, their etiologic role is discussed controversially.8,9

Consistent with this, a recent study investigating the cancer transcriptome of 31 cutaneous SCCs by high-throughput sequencing (Illumina) failed to detect active HPV expression in any of the tumor samples.10 This suggests that HPV may not play role in the continuous proliferation of SCC cells.

Massive parallel sequencing has opened the possibility of comprehensively cataloguing microbial genomes and of analyzing transcriptomes in individual patient samples. Some
high-throughput sequencing platforms (i.e., Illumina or SOLiD) produce short reads but reach a high-sequence coverage and are well suited, e.g., for targeted resequencing of cancer genomes. In contrast, the 454 technology yields fewer reads per run, but the average read length of ~400 bases (with a maximum of up to 1 kb) facilitates similarity searches (e.g., via Basic Local Alignment Search Tool, BLAST). Therefore, this technology is a powerful tool to identify novel viral or bacterial sequences with a sufficient degree of homology to known pathogen genomes. Examples for the discovery of new viruses by this approach include an arenavirus transmitted through solid organ transplantation\(^1\)\(^1\) and the Merkel cell polyomavirus (MCPyV), which was identified in the rare skin tumor Merkel cell carcinoma (MCC) by BLAST-based digital transcriptome subtraction (DTS) out of more than 300,000 pyrosequenced cDNA sequences.\(^1\)\(^2\)

MCPyV as well as two novel human polyomaviruses (HPyVs) identified by degenerate PCR are shed from the skin as assembled virions and therefore seem to belong to the human skin microbiome.\(^1\)\(^3\) The recently discovered trichodysplasia spinulosa-associated polyomavirus appears to be another example for the involvement of novel viruses in skin proliferations of immunocompromised patients.\(^1\)\(^4\) However, the role of MCPyV or other polyomaviruses in NMSC other than MCC is currently unclear: MCPyV DNA has been detected more frequently in NMSCs of immunosuppressed patients, whereas other studies found no evidence for MCPyV infection of NMSCs.\(^8\),\(^1\)\(^5\)–\(^1\)\(^7\)

In our study, we aimed to identify novel viral sequences in cutaneous SCCs from OTR by using 454 high-throughput pyrosequencing. We sequenced cDNA of SCC samples to identify actively transcribed viral sequences and to avoid detection of viral DNA shed from the skin surface. To fully exploit the advantage of the long-read lengths, we used the scoring matrix-based local aligner (BLAST) for DTS.

**Material and Methods**

**Patients and samples**

Tissue biopsies from cutaneous SCCs were obtained from three patients (male 74, female 69 and male 69 years old) who had undergone solid organ transplantation (kidney, double lung or heart–lung transplantation), respectively, 10, 12 or 19 years ago. All patients gave their written informed consent to storage and analysis of the samples, and the study was approved by the local ethics committee. All tumor samples were confirmed as cutaneous SCC by an experienced dermatopathologist, who received the tumor tissue of the resection borders to confirm complete resection of the tumors, while we analyzed the tumor “core.” The HPV-18-positive HeLa cervical cancer cell line served as positive control.

**cDNA library construction and high-throughput sequencing**

From each SCC sample, 50 mg of tumor tissue was lysed in a TissueLyser II (Qiagen, Germany) and total RNA was extracted using Trizol (Invitrogen, Germany). PolyA(+) RNA was extracted using oligo-d(T)-covered magnetic beads (Dynabeads Direct, Invitrogen, Germany) followed by 20-min DNaseI treatment (Promega, Germany) to avoid DNA carryover. After randomly primed reverse transcription (Expand Reverse Transcriptase, Roche, Germany), the second strand was synthesized using Klenow polymerase (New England Biolabs, Germany), and the resulting cDNA was used as the template for Roche 454 Rapid Library procedure (Roche, Germany). Two separate libraries were prepared from sample SSC-1. Integrity and fragment size of the library were evaluated using a 6000 RNA Pico chip (Agilent, Germany) in an Agilent Bioanalyzer.

After amplification by emulsion PCR (GS FLX Titanium SV emPCR Kit, Roche), high-throughput sequencing for each tumor sample was performed on one “Large Region” of a Titanium PicoTiterPlate (for HeLa cells on two 1/8th regions) in a GS FLX Sequencer (Roche, Germany) according to the manufacturer’s instructions.

**Bioinformatics**

To remove artificial and low-quality sequence reads of the obtained 454 data, several filtering steps were applied to the raw data sets. FASTA-corresponding quality files were checked by LUCY software,\(^1\)\(^8\) which excludes low-quality reads (Fig. 1). Low-complexity sequences, defined by the BLAST “dust” filtering option, were removed using in-house Python scripts (available upon request).

To deplete the obtained high-quality (HiFi) data set of human transcripts, we used two methods, DTS\(^1\)\(^9\) with our modifications and the “reference mapping approach.” Both methods used the same human reference databases [RefSeq Human Genomic, Human Transcript, Human assembled chromosomes (GRCh37, HuRef, Hs_Celera) and Human Mitochondrion; see Supporting Information Table S2]. In the DTS approach, data were BLASTed using BLAST\(^{+}\),\(^2\)\(^0\) software package (see sources and parameters in Supporting Information Tables S1 and S3) in two steps: first, XML files obtained using MegaBLAST option were parsed and subtracted for matches using in-house Biopython-based scripts. Second, a BLASTn option was used to identify the remaining reads with matches in human databases. In the alternative way of depleting human transcripts, the “reference mapping approach,” the same HiFi data sets were mapped using the Roche 454 Reference Mapper v2.5.3 to the mentioned human reference databases.

The remaining data of both the DTS and the mapping approach, i.e., reads with no match or not mapped to human sequences, were screened against viral databases by BLASTing at the nucleotide level, initially with default parameters. Non-matching queries were reanalyzed using “low-stringency” parameters. Subsequently, sequences with no match at the nucleotide level were BLASTed at the protein level with BLASTx using default parameters. Finally, sequences without significant matches at both nucleotide and protein levels were checked with the tBLASTx algorithm (Supporting Information Table S3).
All hits to viral database entries were subsequently checked for possible matches in the NR (nonredundant) database using NCBI-BLAST online or EMBL-based WU-BLAST. In addition, for these putative viral hits, the resulting expectancy value, normalized score, alignment length and alignment percent identity were compared with matches obtained from a BLAST search against the NR database. Higher values for the indicated parameters and lower for the expectancy value, respectively, indicated the correct match.

Control sequences
To control the ability of our bioinformatic algorithm to identify reads related to published viral sequences, a part of the large T (LT) antigen of MCPyV coding sequence (positions 1–175; GeneBank NC_010277.1) was manually modified to yield artificial sequences with 93 and 83% homology to the MCPyV LT wt sequence. These sequences were added to the data file as controls. The percent identity between control reads and corresponding database records served as surrogate value for the sensitivity of the analysis.

Results
Results of high-throughput sequencing
454 high-throughput sequencing yielded a total of 717,029 reads for the three SCC samples. On an average, 7.3% of all sequences failed to pass the quality check, leaving 664,822 “HiFi” sequences (a few more reads were filtered out by the reference mapper software previous to mapping; therefore, slightly fewer reads (664,436) were subsequently analyzed for this approach; see Fig. 1). After passing the complete bioinformatic work flow, a total of 662,757 and 662,267 reads matched human sequences using the “DTS” or “reference mapping” approach, respectively. Detailed information for each data set from the three SCC samples is presented in Table 1.

To estimate the coverage of individual cellular genes, we analyzed the number of reads belonging to transcripts expected to have a high abundance in our samples: 6,576 (0.92%) reads covered keratin 14 transcripts (NM_000526.4) expected to be highly abundant in basal keratinocytes. As low-abundant transcripts that were expected to occur only in
subpopulations of cells in our tumor biopsies, we identified nine hits for CD101, three hits for langerin (CD207) and one hit for each CD14 and CD1a, which are markers for Langerhans cells.

Recovery of control sequences
All control sequences (the original and two artificially modified MCPyV LT sequences), which were “spiked” into the data sets before bioinformatic analysis, could be recovered by both ways of data analysis, DTS and reference mapping. At the nucleotide level, the artificially modified control reads resulted in polyomavirus matches only: exact matches to MCPyV and—as the match with the lowest homology—Pan troglodytes verus polyomavirus (HQ385748.1), which was identified with both the 93 and 83% homology control sequences. At the protein level, the hit with the weakest homology recovered was Gorilla gorilla gorilla polyomavirus (ADQ54206.1, 41% identity to the original MPyV control sequence). No species other than polyomaviruses matched to any of the control reads. This confirms the capacity of the bioinformatic approach to identify potentially unknown viruses related to known viruses, and its sufficient level of specificity, even with the lowest stringency of analysis.

Recovery of HPV-18 reads from the HeLa data set
454 sequencing of HeLa cells yielded 82,364 reads. A total of 3.6% were filtered out as low-quality reads and ~96% matched to human databases. Using the DTS and mapping approach, 187 and 192 HPV-18 transcripts, respectively, were identified. Analysis of partially mapped sequences revealed additional 79 sequences, representing fused transcripts of cellular and HPV sequences. For the details, see Supporting Information Table S4.

Characterization of nonhuman hits in SCC samples
A total of 2,171 (DTS approach) and 3,301 (mapping approach) residual sequences were screened for matches with viral and other non-human sequences (Fig. 1). We obtained 11, 7 and 12 viral matches using both DTS and the “mapping” approach for SCC-1, -2 and -3, respectively (Table 1). With one exception (Paramecium bursaria chlorella virus), all hits to viral databases were phages, e.g., Pseudomonas phage or Enterobacteria phage (Table 2; detailed alignments are shown in Supporting Information Fig. S1).

Among the remaining other sequences (i.e., nonviral and non-human), bacterial or fungal hits (e.g., Propionibacteria, Pseudomonas or Streptomyces species as well as Malassezia species; see Supporting Information Table S5) were identified. Moreover, a number of them had nonexact matches to primates and rodents. Because of the low likelihood of mammalian mRNA other than human in our samples, these nonexact matches to primates and rodents were regarded as human. Few reads matching to sequences not expected in the analyzed material, e.g., to plant species or model organisms, such as V. vinifera (grape) or D. rerio (zebrafish) were not

Table 1. Results of 454 high-throughput sequencing of the three cutaneous SCC samples

| Sample  | SCC1            | SCC2            | SCC3            |
|---------|-----------------|-----------------|-----------------|
| Total reads | 205,854         | 311,915         | 199,260         |
| Filtered | 10.88% (22,408) | 2.35% (7,336)   | 11.27% (22,463) |
| Approach | DTS             | Mapping         | DTS             |
| Mapper filter | 0.05% (113)    | None            | 0.05% (113)     |
| Human matches | 88.63% (182,298) | 88.56% (182,298) | 88.56% (182,298) |
| Viral matches | 0.005% (11)  | 0.005% (11)     | 0.005% (11)     |
| Bacterial matches | 0.385% (73) | 0.381% (79) | 0.381% (79) |
| Fungal matches | 0.0019% (4) | 0.001% (6) | 0.001% (6) |
| Other matches | 0.005% (11) | 0.001% (3) | 0.001% (3) |
| Left unassigned | 0.083% (172) | 0.088% (192) | 0.088% (192) |

1Reads matching to sequences not expected in the analyzed material, e.g., to plant species or model organisms. All left, unassigned sequences have been manually checked.
considered as plausible and assigned to the group of “other” matches (Table 1).

Following this analysis, a total of 504 sequences could not be assigned to any entry in any of the databases used (425 of these unassigned sequences could not be assigned by both methods; 37 sequences remained unassigned in the DTS and

42 in the mapping approach only). Each of these sequences was BLASTed without limitation for the expectancy value. The resulting file was analyzed manually, and all putative viral matches or other matches of interest were checked online in EMBL-based WU-BLAST using default parameters (with “dust” filter option off). Although we obtained hits to viral

| Sample S1 | Query ID | Match ID and name | % identities (positives) | E-value | Matches/Alignment length |
|-----------|----------|-------------------|--------------------------|---------|--------------------------|
| 1 | G4LGRVO01BCJ7T | ref|NC_000866.4| Enterobacteria phage T4 | 99 | 0 | 431/434 |
| 2 | G4LGRVO01D77GQ | ref|NC_000866.4| Enterobacteria phage T4 | 99 | 0 | 440/447 |
| 3 | G4LGRVO01DN9V9 | ref|NC_007808.1| Pseudomonas phage PA11 | 99 | 0 | 436/437 |
| 4 | G4LGRVO01AY603 | ref|NC_003309.1| Burkholderia phage phiE125 | 79 | 1E – 60 | 232/297 |
| 5 | G4LGRVO01A8P47 | ref|NC_007623.1| Pseudomonas phage EL | 74 | 3E – 07 | 86/114 |
| 6 | G4LGRVO01EP9FY | ref|NC_012696.1| Roseophage EE36P1 | 71 | 6E – 39 | 261/372 |
| 7 | G4LGRVO01CBFY6 | gb|GU911304.1| Burkholderia phage KL3 | 71 | 2E – 05 | 112/158 |
| 8 | G4LGRVO01DI744 | ref|NC_007870.1| Enterobacteria phage N4 | 67 | 5E – 15 | 278/416 |
| 9 | G4LGRVO01BD8SZ | ref|YP_003358433.1| N4 gp37-like protein | 79 | 1E – 06 | 49/84 |
| 10 | G4LGRVO01A00H | ref|NC_007808.1| Pseudomonas phage PA11 | 54 | 4E – 19 | 239/445 |
| 11 | G4LGRVO01B7JAO | ref|NP_803695.1| ORF129 \[Pseudomonas phage phiKZ\] | 52 | 1E – 06 | 59/115 |

| Sample S2 | Query ID | Match ID and name | % identities (positives) | E-value | Matches/Alignment length |
|-----------|----------|-------------------|--------------------------|---------|--------------------------|
| 1 | GXPDYTX02GJA0X | ref|NC_000866.4| Enterobacteria phage T4 | 99 | 0 | 445/448 |
| 2 | GXPDYTX02IZOOP | ref|NC_009899.1| Paramecium bursaria Chlorrella virus AR158 | 63 | 6E – 07 | 99/159 |
| 3 | GXPDYTX02DFTY | gb|HQ728264.1| Enwinia phage vB_EamMM-Y2 | 61 | 9E – 12 | 118/195 |
| 4 | GXPDYTX02GPJBZ | gb|HQ259105.1| Escherichia phage vB_EcoP_G7C | 56 | 5E – 07 | 124/223 |
| 5 | GXPDYTX02IO6K7 | gb|HQ259105.1| Escherichia phage vB_EcoP_G7C | 56 | 7E – 05 | 125/224 |
| 6 | GXPDYTX02GC99B | ref|YP_003358355.1| N4 gp50-like protein \[Pseudomonas phage LUZ7\] | 53 | 6E – 05 | 46/88 |
| 7 | GXPDYTX02RC9GA | ref|YP_003358439.1| N4 gp33-like protein \[Pseudomonas phage phiKZ\] | 50 | 4E – 11 | 62/124 |

| Sample S3 | Query ID | Match ID and name | % identities (positives) | E-value | Matches/Alignment length |
|-----------|----------|-------------------|--------------------------|---------|--------------------------|
| 1 | G4LGRVO02IU7Y | ref|NC_000866.4| Enterobacteria phage T4 | 100 | 0 | 478/478 |
| 2 | G4LGRVO02IRJW | ref|NC_009231.1| Clostridium phage phiC2 | 80 | 2E – 24 | 116/146 |
| 3 | G4LGRVO02PQ8Q | gb|HM173902.1| Pseudomonas phage PAK_P3 | 75 | 2E – 26 | 156/209 |
| 4 | G4LGRVO02FIPS1 | gb|HM173902.1| Pseudomonas phage PAK_P3 | 75 | 2E – 26 | 156/209 |
| 5 | G4LGRVO02H2BBW | ref|NC_013691.1| Pseudomonas phage LUZ7 | 75 | 5E – 59 | 275/371 |
| 6 | G4LGRVO02ID1W8 | ref|NC_013691.1| Pseudomonas phage LUZ7 | 75 | 8E – 22 | 140/188 |
| 7 | G4LGRVO02H274Y | ref|YP_001956868.1| hypothetical protein 20phi2-1p144 \[Pseudomonas phage 20phi2-1\] | 69 | 8E – 06 | 22/54 |
| 8 | G4LGRVO02D1GQG | ref|NC_013691.1| Pseudomonas phage LUZ7 | 67 | 6E – 06 | 68/103 |
| 9 | G4LGRVO02HC65G | ref|NC_013691.1| Pseudomonas phage phi1 | 60 | 2E – 17 | 175/296 |
| 10 | G4LGRVO02G98S | gb|HQ728264.1| Enwinia phage vB_EamP-S6 | 59 | 2E – 24 | 201/344 |
| 11 | G4LGRVO02HM53C | ref|NC_004456.1| Vibrio phage VHML | 56 | 2E – 21 | 243/439 |
| 12 | G4LGRVO02ICWEG | ref|YP_003358355.1| N4 gp50-like protein \[Pseudomonas phage LUZ7\] | 56 | 2E – 15 | 54/145 |
sequences, these were short, had high expectancy values (i.e., a high chance, that the match was accidental) and they were found among other matches of the same quality. Thus, we concluded that these matches were most likely spurious.

### Discussion

High-throughput sequencing of three cutaneous SCC samples did not reveal known human viral transcripts or any transcripts with sufficient homology to any known human or animal viruses. All control sequences could be recovered indicating that our analysis would have been capable of identifying viral sequences with a low homology to known human or animal viruses (we observed bona fide matches with as little as 41% homology at the protein level). In addition, several viral phage sequences as well as bacterial and fungal hits were identified, which might originate from an environmental source or the skin flora and thus further confirm the validity of the method. Both bioinformatic approaches (DTS or reference mapping) to deplete human sequences from the data set yielded a comparable number of non-human matches for each tumor sample and can therefore be used equally for this purpose.

Our analysis also did not reveal any transcripts of known HPVs or polyomaviruses, which is consistent with the results of others\(^\text{10}\) who did not find papillomavirus mRNA expression in SCCs by transcriptome sequencing using the Illumina technique. Taken together, this might indicate that known viruses such as HPV or HPyV are not transcribed in the majority of tumor cells and may therefore at least not be required for the maintenance of NMSC tumor cells.

Although Feng \textit{et al.} were able to identify a novel polyomavirus, MCPyV, in an analysis of four Merkel cell carcinoma samples\(^\text{12}\) yielding a total of 395,734 sequenced reads (two reads out of these were MCPyV reads), in our study no expressed viral sequences besides phages were found by 454 sequencing of three SCC samples resulting in a total of 717,029 reads. Nevertheless, sequencing of HeLa cells as positive control successfully identified HPV-18 transcripts (0.3% of all reads) with a similar percentage as obtained by others using short-read approaches (i.e., Illumina).\(^\text{10,22}\) The reported levels of tumor virus transcripts in other tumors range between 0.02 and 0.06% for HPV-16-containing cervical and periungual SCCs\(^\text{10}\) and 0.24% for the KSHV-infected tumor cell line BCBL-1.\(^\text{19}\) Given our number of obtained sequence reads from the three SCCs, one viral transcript would have reflected a relative abundance of \(~1–2 \times 10^{-6}\) (1 of 717,029), i.e., much lower than observed for HeLa cells or reported for other tumors.

We conclude that the fact that the 454 technique does not allow sequencing to the same depth as, e.g., the Illumina protocol is balanced by its advantage of a longer read length, thereby providing a better basis for the detection of novel sequences related to known microorganisms. However, the possibility of having missed a viral transcript of very low abundance and/or very low sequence homology to known viruses cannot be completely excluded.

Mapping data can be downloaded using the following links (contact the corresponding author for the password):

- http://www.mh-hannover.de/fileadmin/institute/virologie/downloads/Next-gen_seq_fails_to_Suppl_MappingData.part01.rar
- http://www.mh-hannover.de/fileadmin/institute/virologie/downloads/Next-gen_seq_fails_to_Suppl_MappingData.part02.rar
- http://www.mh-hannover.de/fileadmin/institute/virologie/downloads/Next-gen_seq_fails_to_Suppl_MappingData.part03.rar

### Acknowledgements

The authors thank S. Woltemate for excellent technical assistance and M. Henke for bioinformatics support. T.G. was supported by the DFG SFB 566 and C.H.-G. by the Integrated Research and Treatment Center Transplantation (IFB-Tx) from the German Federal Ministry of Education and Research (BMBF).

### References

1. Euvrard S, Kanitakis J, Claudy A. Skin cancers after organ transplantation. \textit{N Engl J Med} 2003; 348:1681–91.
2. Ulrich C, Kanitakis J, Stockfleth E, \textit{et al.} Skin cancer in organ transplant recipients—where do we stand today? \textit{Am J Transplant} 2008;8:2192–8.
3. Grulich AE, van Leeuwen MT, Falster MO, \textit{et al.} Incidence of cancers in people with HIV/AIDS compared with immunosuppressed transplant recipients: a meta-analysis. \textit{Lancet} 2007;370:59–67.
4. Harteveld MM, Bavinck JN, Kootte AM, \textit{et al.} Incidence of skin cancer after renal transplantation in The Netherlands. \textit{Transplantation} 1990;49:506–9.
5. Adamson R, Obiopo E, Dychter S, \textit{et al.} High incidence and clinical course of aggressive skin cancer in heart transplant patients: a single-center study. \textit{Transplant Proc} 1998;30:1124–6.
6. Euvrard S, Kanitakis J, Poutel-Noble C, \textit{et al.} Aggressive squamous cell carcinomas in organ transplant recipients. \textit{Transplant Proc} 1995;27:1767–8.
7. Martínez JC, Otley CC, Stasko T, \textit{et al.} Defining the clinical course of metastatic skin cancer in organ transplant recipients: a multicenter collaborative study. \textit{Arch Dermatol} 2003;139:301–6.
8. Tuttleton Arron S, Jennings L, Nindl I, \textit{et al.} Viral oncogenesis and its role in nonmelanoma skin cancer. \textit{Br J Dermatol} 2011;164:2102–13.
9. zur Hausen H. Papillomaviruses in the causation of human cancers—a brief historical account. \textit{Virology} 2009;384:260–5.
10. Arron ST, Ruby JG, Dybbro E, \textit{et al.} Transcriptome sequencing demonstrates that human papillomavirus is not active in cutaneous squamous cell carcinoma. \textit{I Invest Dermatol} 2011;131:1745–53.
11. Palacios G, Duce J, Du L, \textit{et al.} A new arenavirus in a cluster of fatal transplant-associated diseases. \textit{N Engl J Med} 2008;358:991–8.
12. Feng H, Shuda M, Chang Y, \textit{et al.} Clonal integration of a polyomavirus in human Merkel cell carcinoma. \textit{Science} 2008;319:1096–100.
13. Schowalter RM, Pastrana DV, Pumphrey KA, \textit{et al.} Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. \textit{Cell Host Microbe} 2010;7:509–15.
14. van der Meijden E, Janssens RW, Lauber C, \textit{et al.} Discovery of a new human polyomavirus associated with trichodysplasia spinulosa in an immunocompromized patient. \textit{PLoS Pathog} 2010;6:509–15.
15. Kassem A, Technau K, Kurz AK, \textit{et al.} Merkel cell polyomavirus sequences are frequently detected in nonmelanoma skin cancer of...
immunosuppressed patients. *Int J Cancer* 2009; 125:356–61.

16. Reisinger DM, Shiffer JD, Cognetta AB, Jr, et al. Lack of evidence for basal or squamous cell carcinoma infection with Merkel cell polymavirus in immunocompetent patients with Merkel cell carcinoma. *J Am Acad Dermatol* 2010;63:400–3.

17. Ridd K, Yu S, Bastian BC. The presence of polymavirus in non-melanoma skin cancer in organ transplant recipients is rare. *J Invest Dermatol* 2009;129:250–2.

18. Chou HH, Holmes MH. DNA sequence quality trimming and vector removal. *Bioinformatics* 2001;17:1093–104.

19. Feng H, Taylor JL, Benos PV, et al. Human transcriptome subtraction by using short sequence tags to search for tumor viruses in conjunctival carcinoma. *J Virol* 2007;81:11332–40.

20. Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. *J Mol Biol* 1990;215:403–10.

21. Altschul SF, Gish W. Local alignment statistics. *Methods Enzymol* 1996;266:460–80.

22. Kostic AD, Ojesina AI, Pedamallu CS, et al. PathSeq: software to identify or discover microbes by deep sequencing of human tissue. *Nat Biotechnol* 2011;29:393–6.