Supplementary Figure 1

**Distribution of mapped reads along the adenovirus genome at 8 and 24 hours post infection.**

Two charts (top and bottom) are shown which show the relative density of sequence reads mapping to the adenovirus genome at 8 and 24 hours post infection. In between the two charts, the figure indicates the relative location of all the adenovirus genes and also highlights the major early regions expected to be significantly expressed at 8 hours post infection. On the bottom chart, the asterisks indicate (from left to right) the location of the transcription start site of the major late transcript, the location of the second and third exons in the obligate tripartite leader sequence present in all major late transcripts. Also note that the height of the chart at 24 hours is some 27 times higher reflecting the greatly increased number of transcripts mapping to the virus genome at that time.
Supplementary Figure 2

**Schematic of the workflow in PIT analysis to generate a Trinity derived proteome.**

This figure illustrates how the data flows through the analysis pipeline to generate the Trinity derived list of peptides and proteins detected based on the assembled transcripts.

1. **Trinity assembled transcripts**
   - Use “getorf” to obtain a list of possible ORFs from the Trinity transcripts
   - Search MS/MS spectra with MaxQuant using the Trinity derived protein list

2. **Map to human genome with GMAP**
   - MaxQuant list of identified peptides and the Trinity transcripts associated with each peptide is reported in “peptides.txt” by MaxQuant
   - Output is a SAM file which describes where each transcript maps to the target genome and the exon structure of each transcript. GMAP loses the FPKM data reported by Trinity so we put it back.

3. **For every transcript in the SAM file add peptides data where it exists in the MaxQuant peptides.txt file. Also add information on ratio changes and confidence score.**
   - Perlscript: pep_to_sam.pl

4. **For each entry in the SAM file with peptides associated with it generate a GFF3 format file entry. Use the reported structure of the transcript in the CIGAR value to describe where on the target genome each MaxQuant identified peptide is derived from. Also report which exon number the peptide starts within and color code each entry to show changes in abundance. Also, for each peptide report the longest possible open reading frame on the transcript by translating the transcript up and down stream of the identified peptide.**
   - Perlscript: pep_to_GFF3_and_orfs.pl

5. **The FASTA file of longest possible ORFs is then checked for duplicate entries and each duplicate is removed whilst concatenating the header information – thus a single FASTA entry may have a header containing information on several peptides which were all derived from that ORF.**
   - Perlscript: batchBlastandParse.pl

**Perlscripts:**
- pep_to_sam.pl
- pep_to_GFF3_and_orfs.pl
- put_fpkm_values_back.pl
- batchBlastandParse.pl

**Supplementary Figure 2**

Nature Methods: doi:10.1038/nmeth.2227
Supplementary Figure 3

**Illustration of integration of the transcriptome and proteome for CHO cells.**

Image taken from the IGV viewer showing a SAM alignment file generated by GMAP using Trinity derived sequences mapped onto the CHO genome. As in figure 1 we show the data from the custom GFF3 file that allows us to see what peptides were identified by MS/MS, their location on the transcript and genome. In this case we are looking at transcript comp4087_c0_seq1 which has 28 distinct peptides associated with one ORF from the transcript. In the BLAST analysis this was identified as being homologous to the human and mouse PREP protein, which is the presequence protease protein present in the mitochondrial membrane which degrades small peptides in a cell. Currently, this gene and protein are listed as “inferred by homology” in the UNIPROT database.

Nature Methods: doi:10.1038/nmeth.2227
### Supplementary Table 4

<table>
<thead>
<tr>
<th></th>
<th>Trinity from T0 dataset</th>
<th>Trinity from T8 dataset</th>
<th>Trinity from T24 dataset</th>
<th>20 million paired end reads</th>
<th>10 million paired end reads</th>
<th>5 million paired end reads</th>
<th>81 million paired end reads</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Peptides identified</strong></td>
<td>27896</td>
<td>26726</td>
<td>20036</td>
<td>27195</td>
<td>25373</td>
<td>21203</td>
<td>28,827</td>
</tr>
<tr>
<td><strong>As a percentage of largest Trinity assembly</strong></td>
<td>97%</td>
<td>93%</td>
<td>70%</td>
<td>94%</td>
<td>88%</td>
<td>74%</td>
<td>100%</td>
</tr>
<tr>
<td><strong>Coverage of protein coding human transcriptome (~30 million bp).</strong></td>
<td>~67x</td>
<td>~60x</td>
<td>~11x</td>
<td>~44x</td>
<td>~22x</td>
<td>~11x</td>
<td>~138x</td>
</tr>
</tbody>
</table>

**Analysis of the effect of reducing sequence coverage on peptide identification.**

The seep sequencing data from each time point was individually used to assemble a transcriptome using Trinity and the reported list of transcripts was used to generate peptide lists as we have described. In addition, the T0 dataset was reduced in size *in silico* by removing increasing amounts of data to produce three data sets of 20, 10 and 5 million paired end reads. Again, each one of these data sets was used to generate a list of transcripts and possible proteins as before. This table compares the outputs from those analyses in terms of the number of peptides identified by MaxQuant showing how reducing the data set size affects the number of peptides reported compared to the data set generated by combining all the data together (a total of 81 million paired end reads). The figures for coverage are approximate and assumes a protein coding transcriptome size of 30 million bases. In addition, not all the sequence reads map to the human genome. So, for example, at T0 approximately 18 million 56bp paired end reads mapped to the human genome which is a coverage of ~67x. This is calculated as 18,000,000 (mapped reads) x 112bp (total length of paired end sequence) / 30,000,000 (approximate size of coding transcriptome) = 67.2.
Supplementary Table 8

Primer 1:
5' ATGGCCAGTCGGGAAGAGGAGGCAGCGCG 3'

Primer 2:
5' TTAGAAATCAAAGGGTTCTGCCGCACGCATC 3'

Primers used to test RNA samples prior to RNAseq for viral RNA
The two primers here (forward and reverse) were designed to amplify a product from DBP mRNA after random primed reverse transcription.