# Lack of correlation between predicted and actual off-target effects of short-interfering RNAs targeting the human papillomavirus type 16 E7 oncogene

J E Hanning[1,3], H K Saini[2,3], M J Murray[1], S van Dongen[2], M P A Davis[2], E M Barker[1], D M Ward[1], C G Scarpini[1], A J Enright[2], M R Pett[1] and N Coleman*,[1]

[1]Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK and [2]EMBL-European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, UK

**Background:** When designing therapeutic short-interfering RNAs (siRNAs), off-target effects (OTEs) are usually predicted by computational quantification of messenger RNAs (mRNAs) that contain matches to the siRNA seed sequence in their 3′ UTRs. It is assumed that the higher the number of predicted transcriptional OTEs, the greater the size of the actual OTE signature and the more detrimental the phenotypic consequences in target-negative cells.

**Methods:** We tested this general assumption by investigating the OTEs of potential therapeutic siRNAs targeting the human papillomavirus (HPV) type-16 E7 oncogene. We studied HPV-negative squamous epithelial cells, from normal cervix (NCx) and skin (HaCaT), which would be vulnerable to 'bystander' OTEs following transfection *in vivo*.

**Results:** We observed no correlation between the number of computationally predicted OTEs and the actual number of seed-dependent OTEs ($P = 0.76$). On average only 20.5% of actual transcriptional OTEs were seed-dependent (i.e., predicted). The unpredicted OTEs included stimulation of innate immune pathways, as well as indirect (downstream) effects of other OTEs, which affected important cancer-associated pathways. Although most significant OTEs observed were seen in both NCx and HaCaT cells, only 0–5.9% of differentially expressed genes overlapped between the two cell types.

**Conclusion:** These data do not support the assumption that actual OTEs correlate well with predicted OTEs.

RNA interference (RNAi) is an endogenous process that causes sequence-specific downregulation of target messenger (mRNA) molecules. It is mediated by short, non-coding RNAs, particularly microRNAs (miRNAs) and short-interfering RNAs (siRNAs). Individual miRNAs generally have numerous mRNA targets (Lim *et al*, 2005), determined via their 'seed-region' (predominantly nucleotides at 5′ positions 2–7; 2–7 nt (Lewis *et al*, 2005)), which binds to a seed-complementary region (SCR) in target mRNAs. Short-interfering RNAs have fewer targets and may produce more specific gene silencing (Okamura and Lai, 2008).

The desired on-target effects of siRNAs are caused through perfect base pairing to the target sequence. In addition, the seed-region of the siRNA antisense (guide) strand can bind SCRs in the 3′ UTR of mRNA molecules that show only partial complementarity, leading to direct seed-dependent effects.

There is considerable interest in exploiting RNAi using exogenous siRNAs for therapeutic targeting of disease-specific genes that are essential for survival of the diseased cells (Phalon *et al*, 2010). Suitable candidates include genes encoded by cancer-associated viruses, as these are not present in uninfected normal

cells. An important example is high-risk human papillomavirus (HR-HPV), a necessary cause of cervical carcinoma (Walboomers *et al*, 1999), as maintenance of HPV oncogene expression is essential for the survival of HPV-positive cervical cancer cells *in vitro* (Magaldi *et al*, 2012).

For RNAi to be therapeutically successful, siRNAs need to be developed and systematically assessed, ensuring they have minimal harmful, unintended 'off-target' effects (OTEs) (Birmingham *et al*, 2007). For some siRNAs, a 2′-*O* methyl ribosyl modification at nucleotide position 2 on the antisense strand can provide a 66% reduction in seed-dependent OTEs (Jackson *et al*, 2006) and a proportionate reduction in seed-independent OTEs that result in a host interferon response (Jackson *et al*, 2006; Judge and MacLachlan, 2008). Introducing DNA into the seed region has also had some success in diminishing OTEs, including when targeting HPV16 oncogenes, but often at the expense of reducing silencing activity (Ui-Tei *et al*, 2008; Yamato *et al*, 2011).

When designing individual siRNAs for therapeutic use, it is important to minimise the possibility of OTEs in cells that do not express the target gene. The approach most commonly used to assess transcriptional OTEs in such cells is to predict computationally the numbers of SCR-containing mRNAs that are distinct from the target gene (Anderson *et al*, 2008). It is generally assumed that the higher the number of such predicted OTEs, the greater the size of the OTE signature and the more likely a detrimental phenotypic effect will be induced (Anderson *et al*, 2008; Vaishnaw *et al*, 2010). In the present study, we used siRNAs against the major HPV16 oncogene E7 as a model system to answer two critical questions. First, do predicted OTEs correlate with actual OTEs? Second, does the cell type studied have an effect on the actual OTEs observed? We compared the numbers of computationally predicted OTEs for four different siRNAs designed to target HPV16 E7 with the actual OTEs observed following the treatment of HPV-negative keratinoctyes (squamous epithelial cells). Such cells, in which the target gene was entirely absent, provided a particularly useful system for detailed study of the phenotypic and transcriptional OTEs that would occur in 'bystander' non HPV-infected keratinocytes adjacent to cervical carcinoma cells. We focussed on siRNAs with the 2′-*O* methyl ribosyl modification, as this is most likely to be included in future therapeutic siRNAs (Vaishnaw *et al*, 2010). Our work has identified significant limitations in current OTE prediction algorithms, indicating a continuing requirement for new approaches to OTE identification.

## MATERIALS AND METHODS

**Cell culture.** The squamous epithelial cell lines used were CaSki, derived from an HPV16-positive human cervical squamous cell carcinoma (Pattillo *et al*, 1977), and HaCaT, an immortalised HPV-negative epidermal (skin) cell line (Boukamp *et al*, 1988). The cells were authenticated by short tandem repeat profiling by the American Type Culture Collection, Manassas, VA, USA. Primary cultures of normal cervical keratinocytes (NCx) were obtained from hysterectomy specimens removed for non-neoplastic disease unrelated to the cervix. The NCx cells were HPV negative by nested PCR and reverse-line blot hybridisation (Muralidhar *et al*, 2007). Cell culture and determination of growth rates were performed as described (Pett *et al*, 2004; Herdman *et al*, 2006). Briefly, cells were stained with 0.4% trypan blue (Life Technologies Inc., Grand Island, NY, USA) and viable cells counted using a Countess automated cell counter (Life Technologies Inc.).

**siRNA design.** Seven siRNAs targeting the HPV16 oncogene E7 were chosen for initial testing. These represented four previously published siRNAs, E7-Tang (Tang *et al*, 2006), E7-Jiang (Jiang and Milner, 2002), E7–573 (Yamato *et al*, 2008) and E7–752 (Yamato *et al*, 2008), plus three novel siRNAs designed using Dharmacon siDESIGN centre (Thermo Scientific, Waltham, MA, USA). All siRNAs were manufactured by Dharmacon (Birmingham *et al*, 2007) (Thermo Scientific), and all contained the 2′-*O*-methyl ribosyl modification at position 2 on the antisense strand (Jackson *et al*, 2006). Of the novel siRNAs, E7–653 and E7–141 were selected as they had the highest thermodynamic score and were thus predicted to give the greatest depletion of E7. E7–127 was chosen as it had the highest thermodynamic score of the siRNAs that were classed as having 'low-seed frequency' (Table 1). Predicted OTEs were determined according to criteria used by Dharmacon – namely the number of mRNA transcripts in the whole transcriptome containing an SCR in the 3′ UTR that corresponded to 5′ 2–7 nt on the siRNA antisense strand. Such SCRs were identified by searching a database of 3′ UTRs obtained from *Sylarray* (Bartonicek and Enright, 2010), on the basis of *Ensembl v57* (http://www.ensembl.org/index.html).

**siRNA transfection.** Cells were cultured in six-well plates, with CaSki and HaCaT seeded at $1.6 \times 10^5$ cells/well and NCx at $1.9 \times 10^5$ cells/well. Cells were transfected at 30–40% confluency, following removal of fibroblast feeder cells using $1 \times$ phosphate-buffered saline ($1 \times$ PBS) in the case of NCx. In parallel with each E7 siRNA experiment, cyclophilin B depletion was performed using cyclophilin B siRNA (Thermo Scientific), as a transfection efficiency indicator. Further transfection efficiency analysis was carried out using a fluorescently labelled siRNA (siGLO green transfection indicator, Thermo Scientific). All transfections used lipofectamine RNAiMAX (Invitrogen, Paisley, UK) at 9.6 $\mu$l per well. The final siRNA concentrations used were the lowest that provided highest transfection efficiency: these were 37 nM unless stated otherwise. For NCx, feeder cells were added back 12 h after transfection. Medium was replaced with standard medium at 24 h post-transfection and replaced every 24 h thereafter, until cells were harvested. We also used a pool of non-targeting control siRNAs (ON-TARGET plus non-targeting, Thermo Scientific), at concentrations identical to those of the test E7-targeting siRNAs.

| Table 1. Initial analysis of siRNAs targeting HPV16 E7 | | | | | |
|---|---|---|---|---|---|
| siRNA source | Name | siRNA sense (passenger) strand sequence | No. 3′ UTR hits[a] | +/+ max hit (nt)[b] | Thermodynamic score[c] |
| Tang (19) | E7-Tang | 5′-GCACACACGUAGACAUUCG-3′ | 4255 | 15 | 65 |
| Dharmacon | E7–127 | 5′-GGACAAGCAGAACCGGACA-3′ | 846 | 14 | 73 |
| Dharmacon | E7–141 | 5′-GGACAGAGCCCAUUACAAU-3′ | 4501 | 15 | 88 |
| Dharmacon | E7–653 | 5′-GCUCAGAGGAGGAGGAUGA-3′ | 4640 | 18 | 95 |

Abbreviations: HPV = human papillomavirus; siRNAs = short-interfering RNAs.
[a]Number of predicted OTEs.
[b]Highest number of consecutive nucleotide matches between the siRNA and the human genome.
[c]Thermodynamic score showing the depletion potential of the siRNA, on the basis of several parameters (7).

**RNA extraction and quantitative reverse transcription PCR (qRT-PCR).** RNA was harvested using 1 ml of TRIzol (Life Technologies) per well, after feeder cell removal in the case of NCx. RNA quality was assessed using the Agilent Bioanalyzer 6000 (Agilent, Stockport, UK) and the RNA stored at −80 °C until used. Levels of RNA depletion were measured by qRT-PCR. QuantiTect reverse transcription kit (Qiagen, Crawley, UK) was used for cDNA synthesis, using polyT primers alone (Qiagen). Quantitative PCR was performed using SYBR Green JumpStart Taq ReadyMix (Sigma-Aldrich, Dorset, UK), with primers (Sigma-Aldrich, unless specified) listed in Supplementary Table S1. Expression levels were normalised against the mean of three house-keeping genes (*ACTB*, *GAPDH* and *TBP*) (Herdman *et al*, 2006). Quantification of miR-137 and miR-302f levels was performed using TaqMan qRT-PCR (Applied Biosystems, Warrington, UK), as described (Muralidhar *et al*, 2007).

**Western immunoblotting.** Cells were harvested with RIPA buffer (Sigma-Aldrich) containing protease inhibitor (Roche, Basel, Switzerland), following the removal of feeder cells in the case of NCx. Extracts were stored in 20% glycerol at −80 °C. Total-protein concentration was determined using BCA (bicinchoninic acid) Protein Assay Kit (Thermo Scientific). Seventy micrograms of total protein was run on a 12% pre-cast NuPage polyacrylamide gel (Invitrogen), blotted and analysed by hybridisation of antibodies against HPV16 E7 (ED17, Santa Cruz Biotechnology, CA, USA) or HPV16 E6 (1E–6F4, Euromedex, Souffelweyersheim, France). Detection was carried out using ECL-Plus Western Blotting Detection System (GE Healthcare, Amersham, UK). Protein levels were normalised to those of $\beta$-tubulin (ab6046, Abcam, Cambridge, UK).

**Microarray expression profiling.** Total RNA (1 $\mu$g) was labelled and hybridised as described (Winder *et al*, 2011). The bead-level data (raw text files containing background-corrected foreground intensities) obtained from BeadScan software were read using the Bioconductor *beadarray* package (version 2.2.0) (Dunning *et al*, 2008). The data were converted to bead-summary data using *summarise* function and the default *logGreenChannelTransform* transformation (Dunning *et al*, 2007), after detecting and removing the spatial artefacts using the BeadArray Subversion of Harshlight (BASH) tool, as part of the Bioconductor *beadarray* package (Barbosa-Morais *et al*, 2010). The mean and s.d. of the remaining beads were calculated. The log$_2$-transformed data were then quantile normalised. The probes were filtered on the basis of the expression-detection score ($P < 0.05$) and duplicate probes further removed on the basis of the interquartile range filtering. We used the Bioconductor annotation file *illuminaHumanv4.db* (Barbosa-Morais *et al*, 2010) to retrieve Entrez Gene identifiers and gene symbols for the probes. The raw (.txt) and normalised intensities files are available at ArrayExpress (*Accession*: E-MTAB-967).

To detect differentially expressed genes, the normalised expression values were analysed using linear models with *limma* (Smyth, 2004). Significance of differential expression was tested by an empirical Bayes moderated *t*-test and adjusted for multiple-testing using Benjamini and Hochberg's method (Benjamini *et al*, 2001). For each siRNA, we obtained lists of differentially expressed probes with more than two-fold change (log$_2$ fold-change of 1) in either direction, and an adjusted *P*-value of < 0.05, compared with untreated cells. The overlap between genes differentially expressed in NCx and HaCaT cells was plotted using the Bioconductor *venndiagram* package (http://faculty.ucr.edu/∼tgirke/Documents/R_BioCond/My_R_Scripts/vennDia.R).

**Pathway enrichment analysis.** The lists of up- and downregulated genes were used for Gene Ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway analysis, performed using the conditional *hyperGTest* function from the Bioconductor

*GOstats* package (Falcon and Gentleman, 2007). To obtain the gene universe, we removed probes without Entrez Gene identifiers. The lists of differentially expressed genes were evaluated for the enrichment of GO 'Biological Process' terms and KEGG pathways using mappings defined in the Bioconductor *illuminaHumanv4.db* annotation library (Barbosa-Morais *et al*, 2010). Categories with adjusted *P*-values < 0.05 (Benjamini *et al*, 2001) were considered significant.

**Gene-set enrichment analysis and functional maps.** We also analysed genelists using gene-set enrichment analysis (GSEA) to identify statistically over-represented groups of genes on the basis of the expression data (Subramanian *et al*, 2005). The gene sets used (http://baderlab.org/GeneSetDB_02?action=AttachFile&do=view&target=GO_K_NCI_BIOC_PF_Hs_eg.GMT) were derived from: GO (Biological Process, Molecular Function and Cellular Component), NCI pathways, KEGG pathways, Reactome and PFAM protein domain families. Enriched gene sets were graphically organised into a network using the Cytoscape (Shannon *et al*, 2003) plug-in *Enrichment Map* (Merico *et al*, 2010), where each gene set was represented as a node and edges represented overlap between sets. Node size was proportional to the total number of genes belonging to the corresponding gene set, while edge thickness was proportional to the overlap score. We used enriched gene sets with false-discovery rate *q*-value < 5% to build the enrichment map.

**Sylamer algorithm.** Sylamer assesses for enrichment and/or depletion of nucleotide words of specific length, complementary to elements of the seed region of miRNAs/siRNAs, in the 3′ UTRs of genes within ranked lists (van Dongen *et al*, 2008). Significance is calculated using hypergeometric statistics. We derived a single-summed significance score for each seed-dependent OTE, as described previously for miRNA effects (Palmer *et al*, 2010). This score was produced by combining Sylamer results for the set of the core 2–7 nt hexamer, the two heptamers and the single octamer particular to each enriched SCR (complementary to siRNA 1–8 nt). The resulting scores fitted an extreme value distribution, allowing *P*-values to be assigned, with values < 0.01 being deemed significant (Palmer *et al*, 2010). We also used a simple $\chi^2$ contingency test to confirm the significance of any identified SCR enrichment in downregulated genes for each siRNA.

## RESULTS

**siRNA potency and predicted OTEs.** For each of the seven siRNAs, we first determined their potency, defined as the ability to deplete target HPV16 E7 mRNA and protein in CaSki cells. As E7 is co-transcribed with E6 in polycistronic transcripts (Smotkin and Wettstein, 1986), we measured depletion of both E7 and E6 at the RNA and protein level, using qRT-PCR and western blotting, respectively. We observed minimal fluctuation in transfection efficiency across control experiments, with mean cyclophilin B depletion of 97% (Supplementary Figure S1A) and mean siGLO transfection of 97% (Supplementary Figure S1B).

Of the four previously published siRNAs (E7-Tang, E7-Jiang, E7–573 and E7–752), only E7-Tang had a potency of > 70% (Figure 1A). All three of the novel E7-targeting siRNAs (E7–653, E7–141 and E7–127) were highly potent at depleting E7 and E6 RNA and protein, particularly at 48 h, which was therefore chosen as the time point to assess OTEs (Figure 1B). On the basis of the potency data, OTE analysis was only performed on the three novel siRNAs and E7-Tang.

By searching for SCRs in 3′ UTRs, we determined that all the top three, most potent siRNAs (E7–141, E7-Tang and E7–653), had a similar number of predicted OTEs. In contrast, E7–127 had over five-fold fewer predicted OTEs (Table 1). If the assumptions

about the prediction algorithm were correct, it followed that E7–127 would induce the fewest significant OTEs *in vitro*.

**Phenotypic OTEs.** We first measured growth inhibitory effects of the four selected siRNAs over a 6-day period, following the transfection of HaCaT cells. As HaCaT is an HPV-negative cell line, any observed phenotypic changes were due to siRNA OTEs. We were unable to use normal cervical keratinocytes (NCx) for this purpose, as growth of NCx cells is sensitive to a range of factors, including confluency and the extent of fibroblast feeder cell support, making direct comparisons of growth rate difficult. In contrast, HaCaT cells do not require feeder cell support and provide a straightforward system for quantifying cell growth.

Although the most potent siRNA, E7–141, had no effect on growth, each of the other three siRNAs caused growth inhibition (up to 28% compared with control-treated cells) (Figure 2). These effects were not associated with differences in numbers of predicted OTEs, as E7–141 had the second highest number of predicted OTEs (Table 1). We confirmed that none of the siRNAs contained 5′ UGGC motifs, which can induce toxic cellular effects (Fedorov *et al*, 2006).

In view of the unpredicted effects on cell growth, we next undertook a detailed analysis of the transcriptional OTEs of all four siRNAs. We used a microarray approach to quantify the changes in the global transcriptome of HPV-negative keratinocytes, studying both primary cultures of NCx and HaCaT (skin) cells, with all experiments performed in biological triplicate. In these experiments, transfection efficiency was 87.1–91.8% for NCx and 90.7–93.7% for HaCaT.

**Actual *vs* predicted numbers of transcriptional OTEs.** We determined actual transcriptional OTEs by comparing gene expression profiles in siRNA-treated cells *vs* untreated controls, and identifying differentially expressed transcripts. To assess any non-specific OTEs, we looked for genes that were significantly differentially expressed (upregulated or downregulated), following treatment with all of the four siRNAs (Figure 3A). There were no such genes in HaCaT and only one gene in NCx (*AGR2*). We concluded that the transfection reagent and nonspecific siRNA effects did not cause any significant OTEs and therefore did not interfere with our downstream analysis.

Actual seed-dependent OTEs were defined as the number of downregulated transcripts with a significant two-fold change (log2FC $< = -1$ and adjPval $< 0.05$) that also contained an SCR (corresponding to 2–7 nt on the 5′ siRNA antisense strand) in their 3′ UTR. Across all the eight experiments (four siRNAs, each in two cell lines), there was no correlation between the numbers of predicted and actual seed-dependent transcriptional OTEs (*P*-value $= 0.76$) (Figure 3B). Indeed, the two siRNAs with the largest numbers of predicted OTEs (E7–141 and E7–653) caused downregulation of approximately the same number of genes as E7–127 (Figure 3B, Table 2), which had five-fold fewer predicted OTEs (Table 1).

The siRNAs were categorised into groups with high- and low-predicted SCR frequency, according to the previously used criteria (Anderson *et al*, 2008). By this process, there were three siRNAs with 'high' numbers of predicted SCRs ($> 4000$) and one with 'low' numbers ($< 1000$). There was no significant difference in actual OTEs between the two groups (*P*-value 0.65: Figure 3C).
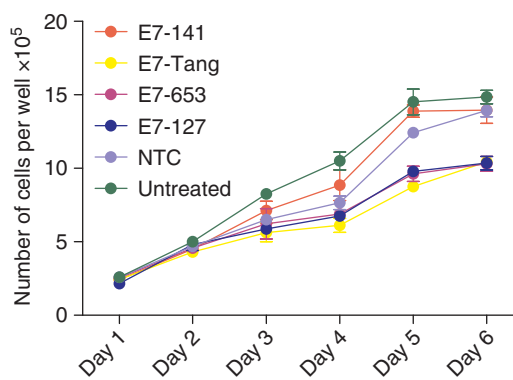


**Figure 2.** Assessment of phenotypic siRNA OTEs *in vitro*. Growth of HPV-negative HaCaT cells was quantified for 6 days post-transfection of test siRNAs, compared with untreated cells (untreat) and cells treated with non-targeting control (NTC) siRNA. Data were obtained from two experimental replicates for E7-Tang, E7–127 and E7–653 and from four experimental replicates for E7–141, NTC and untreated cells.
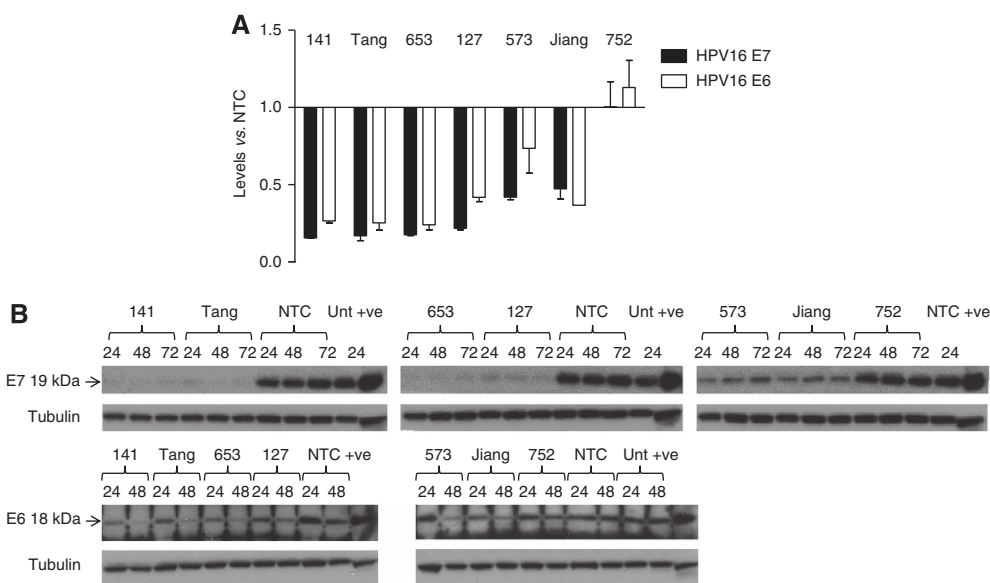


**Figure 1. Potency of all seven siRNAs tested. (A)** Depletion of HPV16 E7 and E6 mRNA in CaSki cells treated with each siRNA, compared with cells treated with non-targeting control (NTC) siRNAs. The E6 primers detected full-length E6 transcripts. Error bars represent the s.e. from three experimental replicates. **(B)** Effects on protein levels. The upper panel shows levels of HPV16 E7 protein at 24, 48 and 72 h following transfection, whereas the lower panel shows levels of E6 protein at 24 and 48 h. Tubulin-loading control is shown underneath each blot.
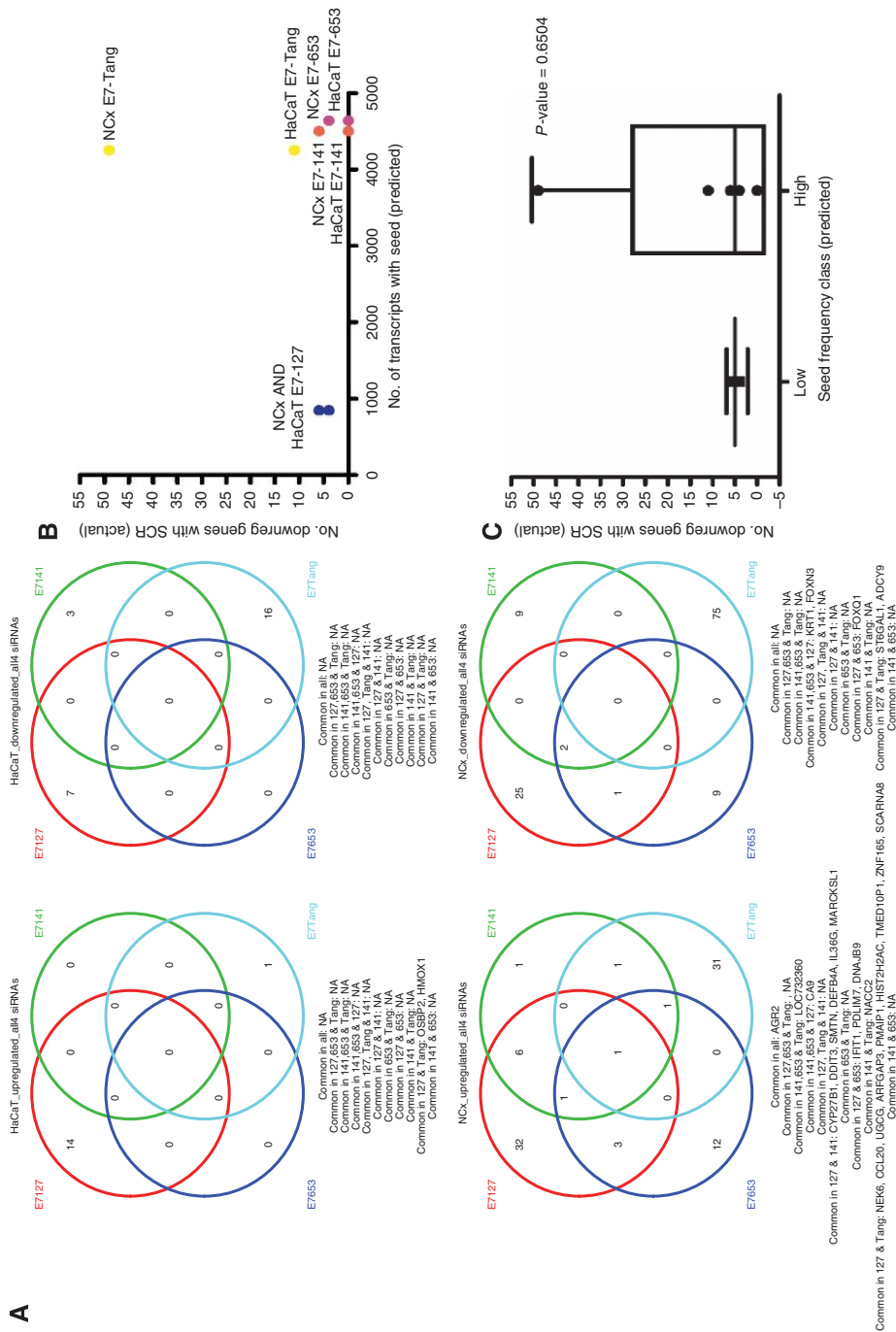
Figure 3. **Numbers of transcriptional OTEs. (A)** Venn diagrams showing the overlap of genes differentially expressed, following the siRNA treatment in HaCaT (top row) and NCx (bottom row). The left column shows upregulated genes and the right column shows downregulated genes. **(B)** Comparison of predicted and actual seed-dependent OTEs for four siRNAs in two different cell types. By two-tailed linear regression analysis, $P = 0.76$. **(C)** Actual seed-dependent OTEs in siRNAs with low and high numbers of predicted OTEs (bar = median, box = IQR, whiskers = range). By unpaired *t*-test there was no significant difference between the two groups (*P*-value 0.65).

| Table 2. Summary of actual transcriptional OTEs observed | | | | | | |
|---|---|---|---|---|---|---|
| siRNA (No. 3′ UTR hits) | Cell line | Total no. of differentially expressed transcripts | No. of downregulated transcripts | No. of downregulated transcripts with SCR (%) | No. of upregulated transcripts | No. of upregulated transcripts with SCR (%) |
| E7–127 (846) | HaCaT | 23 | 7 | 4 (17.4) | 16 | 0 |
| | NCx | 83 | 30 | 6 (7.2) | 53 | 0 |
| E7-Tang (4255) | HaCaT | 19 | 16 | 11 (57.9) | 3 | 0 |
| | NCx | 121 | 77 | 49 (40.5) | 44 | 9 (7.4) |
| E7–141 (4501) | HaCaT | 3 | 3 | 0 | 0 | 0 |
| | NCx | 22 | 11 | 6 (27.3) | 11 | 2 (9.1) |
| E7–653 (4640) | HaCaT | 0 | 0 | 0 | 0 | 0 |
| | NCx | 30 | 12 | 4 (13.3) | 18 | 6 (20) |

Abbreviations: OTEs = off-target effects; SCR = seed-complementary region; siRNAs = short-interfering RNAs.
Probes were considered differentially expressed if their $\log_2$FC was $\geqslant 1$ or $\leqslant -1$, with an adjusted *P*-value <0.05.

Consistent with previous reports (Grimson *et al*, 2007; Hammell *et al*, 2008; Selbach *et al*, 2008), we observed that 1.15% or less of mRNAs that contained an siRNA SCR in their 3′ UTR were significantly downregulated (range 0–1.15%). Moreover, for each analysis, the actual seed-dependent OTEs represented on average only 20.5% of all differentially-expressed genes (range 0.0–57.9%) (Table 2, Figure 4). Using the $\chi^2$–test, we found significant enrichment in downregulated genes ($P<0.01$) of SCRs for E7–127 and E7-Tang, in both HaCaT and NCx (Supplementary Table S2A).

**Sylamer analysis of seed-dependent OTEs.** We used Sylamer to test for enrichment and/or depletion of all possible 6, 7 and 8-mer words in cells treated with each of the four E7-targetting siRNAs, compared with untreated cells. Two of the siRNAs, E7–127 and E7-Tang, were found to induce significant actual siRNA seed-dependent OTEs. These were seen in both HaCaT and NCx (Figure 5A). The seed-dependent OTEs of E7-Tang were more highly significant than those of E7–127 (Supplementary Table S2B). Interestingly, we observed that E7-Tang shared 3–8 nt of its antisense strand with the seed region (2–7 nt) of the endogenous miRNA miR-1, resulting in significant downregulation of mRNAs containing miR-1 SCRs. In contrast, none of the other siRNAs showed a match for endogenous miRNA seeds. Non-targeting control siRNA did not induce significant seed-dependent OTEs (Supplementary Figure S2A).

We also identified that E7–653 caused a sequence-dependent OTE in NCx (but not in HaCaT) that was not due to the E7–653 seed sequence. The OTE was due to mRNAs in which the 6-mer nucleotide sequence 'AGCAAT' was enriched in the 3′ UTR (*P*-value 1e − 11.3) (Figure 5B). This sequence was complementary to 3–8 nt of the seed region of miR-137 and miR-302f, but not the 2–7 nt core seed region of these miRNAs. $\chi^2$ assessment showed no significant enrichment of the 2–7 nt SCR of these miRNAs in E7–653-treated HaCaT and NCx (Supplementary Table S2A). Moreover, using Taqman qRT-PCR, we found no elevation in levels of either miRNA following E7–653 transfection of NCx or HaCaT (data not shown). Interestingly, E7–653 transfection of NCx (but not HaCaT) also produced significant enrichment in downregulated genes of the longer 8-mer SCR sequence 5′-CAGC AATA-3′ (*P-value* 1e − 6.1), and both the constituent 7-mers 5′-CAGCAAT-3′ (*P*-value 1e − 10.4) and 5′-AGCAATA-3′ (*P-value* 1e − 8.4), each of which contained the 5′-AGCAAT-3′ 6-mer (Figure 5B).
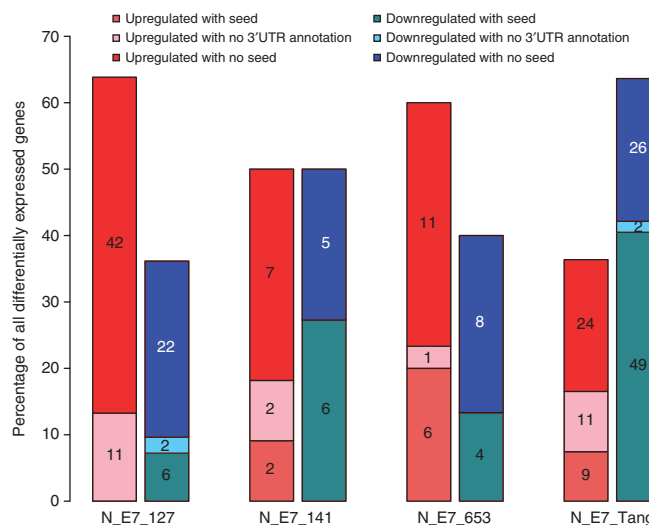


**Figure 4.** Features of the observed transcriptional OTEs in NCx cells. The bars depict categories of mRNA transcripts differentially expressed following siRNA treatment. For each pair, the left-hand bar depicts the upregulated genes and the right hand bar depicts the downregulated genes.

**Enrichment of endogenous miRNA SCRs.** As siRNA transfection can upregulate endogenous miRNA targets (Jackson and Linsley, 2010), we determined the frequency of the 2–7 nt SCR for a set of 345 conserved human miRNAs in mRNAs upregulated following treatment with each siRNA, compared with the frequency in non-upregulated genes. On average, 22.5% of upregulated genes contained one or more miRNA SCR, which was comparable to the frequency in the non-upregulated genes (Supplementary Table S3). In addition, Sylamer assessment of all possible 6-, 7- or 8-mer words showed no significant enrichment of endogenous miRNA SCRs in upregulated genes for any of the four siRNAs in either HaCaT or NCx (Figure 5A). Thus, there was no general enrichment of miRNA SCRs in genes upregulated following siRNA treatment.

**Pathway analysis of OTEs.** To determine whether there was an association between the numbers of OTEs induced and significance at a pathway level, we performed pathway analysis of differentially
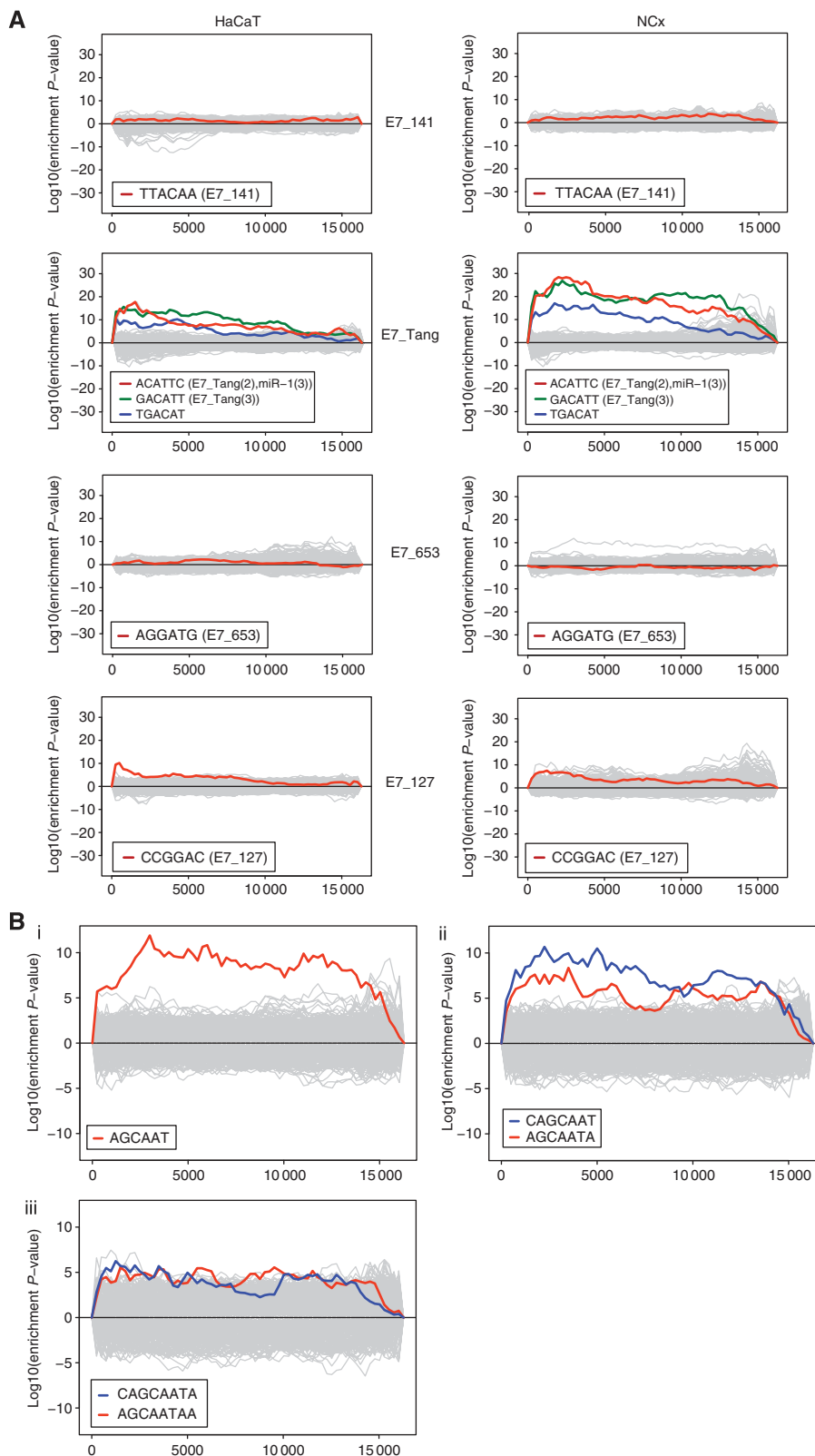
**Figure 5.** Sylamer analysis of OTEs. Each graph is a Sylamer landscape plot. The x-axis shows the ranked genelist, with genes downregulated following siRNA transfection to the left and upregulated genes to the right. The y-axis shows $\log_{10}$-transformed and sign-adjusted enrichment *P*-values for each SCR word, relative to *P*-values of all other words. (**A**) Seed-dependent OTEs of E7–141 (row (i)), E7-Tang (row (ii)), E7–653 (row (iii)) and E7–127 (row (iv)) in HaCaT (left column) and NCx cells (right column). In all cases, treated cells were compared with their untreated counterparts. For each siRNA, the 6-mer sequences complementary to the 2–7 nt seed-regions are shown in red. Any other sequences derived from the region around the 6 mers that reached statistical significance are shown in green and blue. In row (ii), E7-Tang(2) and miR-1(3) denote the sequence complementary to 2–7 nt in E7-Tang and 3–8 nt in miR-1, whereas E7_Tang(3) denotes the sequence complementary to 3–8 nt in E7-Tang. (**B**) Other sequence-dependent OTEs produced by E7–653 in NCx cells. The panels show non-SCR sequences that reached statistical significance, with 6 mer in (i), 7-mers in (ii) and 8-mer in (iii).

expressed genes using GO, KEGG and GSEA pathway analysis. We applied strict criteria in this exercise. We required pathways to be present in at least two of the three analyses to be considered enriched and used cut-offs of adjusted (rather than unadjusted) *P*-values <0.05 for GO and KEGG and false discovery rate <0.05 for GSEA. Detailed results are given in Supplementary Tables S4–S23.

For E7–141, no pathway was enriched in more than one analysis in either cell line, for either the up- or downregulated genes. Transcripts downregulated by E7-Tang were linked to reduced DNA replication, through several pathways (Figure 6A). In addition, in NCx cells, transcripts that were downregulated by E7-Tang and contained the corresponding SCR showed enrichment for the term 'regulation of actin cytoskeleton'. As miR-1 is functionally linked to actin cytoskeleton organisation (Jiang *et al*, 2010), the latter observation suggested that an siRNA containing an endogenous miRNA seed region was capable of acting similarly to the endogenous miRNA.

In both NCx and HaCaT, E7–127 and E7–653 induced significant upregulation of genes related to innate immune responses (Figure 6B and C). E7–127 also showed enrichment for cell cycle pathways, such as DNA replication and cell division (Figure 6B), in downregulated genes. Such genes were not enriched in the SCR-containing downregulated genes, suggesting that the effect was more likely to be because of the upregulation of innate immune response genes seen following E7–127 treatment (Olejniczak *et al*, 2011). Although GSEA demonstrated upregulation of immune response genes in NCx (but not HaCat) following treatment with non-targeting control siRNA (Supplementary Figure S2B), no significant pathways were identified by GO or KEGG.

**Comparison of OTEs in NCx *vs* HaCaT.** In view of the different sequence-dependent OTEs induced by E7–653 in NCx and HaCaT, we compared all the transcriptional OTEs induced by the four siRNAs in the two cell types. Although all the significant seed-dependent and immune-related OTEs that we observed were seen in both NCx and HaCaT, there was little overlap in the actual transcripts that showed differential expression in the two cell types (Supplementary Figure S3A). Of the total number of differentially expressed transcripts for each siRNA, the percentage that was differentially expressed in both NCx and HaCaT was only 0–5.9% (dependent on the stringency of the cut-off used) (Supplementary Figure S3A and B). These low values were not associated with major differences in the baseline transcriptional profiles of NCx and HaCaT, as only 8.6% of expressed genes showed significant differential expression between untreated NCx and HaCaT cells (Supplementary Figure S4).

## DISCUSSION

Current computational algorithms predict siRNA-induced OTEs by calculating the number of transcripts containing a 3′ SCR corresponding to the siRNA seed region. It is generally assumed that siRNAs with a greater number of predicted 'hits' will produce a greater number of actual transcriptional OTEs and hence a more deleterious phenotype (Anderson *et al*, 2008). Because of this, current pipelines for therapeutic siRNA development exclude siRNAs with high number of predicted OTEs wherever possible (Vaishnaw *et al*, 2010). Our data indicate that the underlying assumption is flawed, as we found no correlation between the numbers of predicted and actual seed-dependent OTEs for potential therapeutic siRNAs targeting HPV16-E7.

Review of published literature provides evidence to support our findings. It was previously shown that there was a positive correlation between the number of predicted OTEs and the number of actual OTEs in siRNAs classed as having 'low' (<350)

and 'medium' (∼2500–2800) numbers of predicted SCR hits (Anderson *et al*, 2008). However, there was no clear difference in actual OTEs between siRNAs classed as having 'low' and 'high' (>3800) numbers of predicted OTE hits (Anderson *et al*, 2008). Indeed, in the present study, we observed that the siRNAs with high numbers of SCRs (E7-Tang, E7–141 & E7–653) had no more actual seed-dependent OTEs than the siRNA with low numbers of SCRs (E7–127).

Our data also indicated that the consequences of the OTEs induced by therapeutic siRNAs are not related to the number of actual OTEs but to the nature of the genes affected. While two of the HPV16 E7 siRNAs tested (E7–127 and E7–141) induced similar numbers of actual OTEs, only E7–127 significantly modulated cellular pathways and induced significant seed-dependent OTEs on Sylamer analysis. There was an association between phenotypic OTEs *in vitro* and the presence of significant transcriptional OTEs, as the three siRNAs that reduced the growth of HaCaT cells (E7-Tang, E7–653 and E7–127) all produced significant transcriptional OTEs that consistently affected key cellular pathways. In contrast, E7–141 did not induce significant transcriptional OTEs and did not affect cell phenotype. Differences in the induction of significant transcriptional OTEs were not predicted by current bioinformatic algorithms, nor by the total number of OTEs induced. Instead, pathway analysis was required for them to be identified.

It is interesting that the two siRNAs with significant seed-dependent OTEs on Sylamer analysis (E7–127 and E7-Tang) had the lowest number of predicted OTEs. These findings support the notion that a general negative correlation may exist between the number of available SCR-containing genes (i.e., predicted OTEs) and actual OTEs, as a greater range of mRNA targets bound by an siRNA may reduce the capacity of that siRNA to deplete levels of individual mRNAs significantly (Arvey *et al*, 2010). This view is consistent with the evidence that pooling siRNAs reduces the number of significant actual OTEs, despite increasing the number of predicted OTEs (Kittler *et al*, 2007). Reduced siRNA concentration is unlikely to be the only explanation for such an effect, as lowering the concentration of one siRNA in a pool did not reduce the number of actual OTEs (Kittler *et al*, 2007). In contrast, our data indicated that large numbers of predicted OTEs do not necessarily reduce siRNA potency (Arvey *et al*, 2010). The potency of the four siRNAs tested corresponded more closely to their thermodynamic score (Birmingham *et al*, 2007) than to the numbers of predicted OTEs.

Of the actual transcriptional OTEs induced by the four siRNAs, on average only 20.5% were due to direct seed-dependent OTEs. In view of this, we tested for other types of OTE computationally (Supplementary Figure S5). We looked for evidence of modulation of endogenous RNAi pathways, leading to upregulation of targets of endogenous miRNAs. There was no general enrichment of miRNA SCRs in upregulated genes compared with non-upregulated genes, and no SCRs were enriched on Sylamer analysis, indicating that endogenous miRNA imbalances did not induce significant shifts in overall mRNA profiles. However, Sylamer may have underestimated miRNA effects. The presence of a minimal 6-mer seed match may be insufficient to detect miRNA targets, as miRNA regulation is influenced by additional targeting parameters and also depends on the expression levels of miRNAs in the system studied (Bartel, 2009).

We did observe evidence of seed-independent OTEs, which are thought to be caused by siRNAs binding intracellular receptors (Jackson and Linsley, 2010). Two of the siRNAs tested, E7–127 and E7–653, stimulated innate immune pathways in both NCx and HaCaT. Interestingly, E7–127 had the lowest number of predicted OTEs, which may have resulted in the highest amount of unbound siRNAs capable of stimulating host dsRNA receptors, such as RIG-I-like and NOD-like receptors (Sioud, 2010).
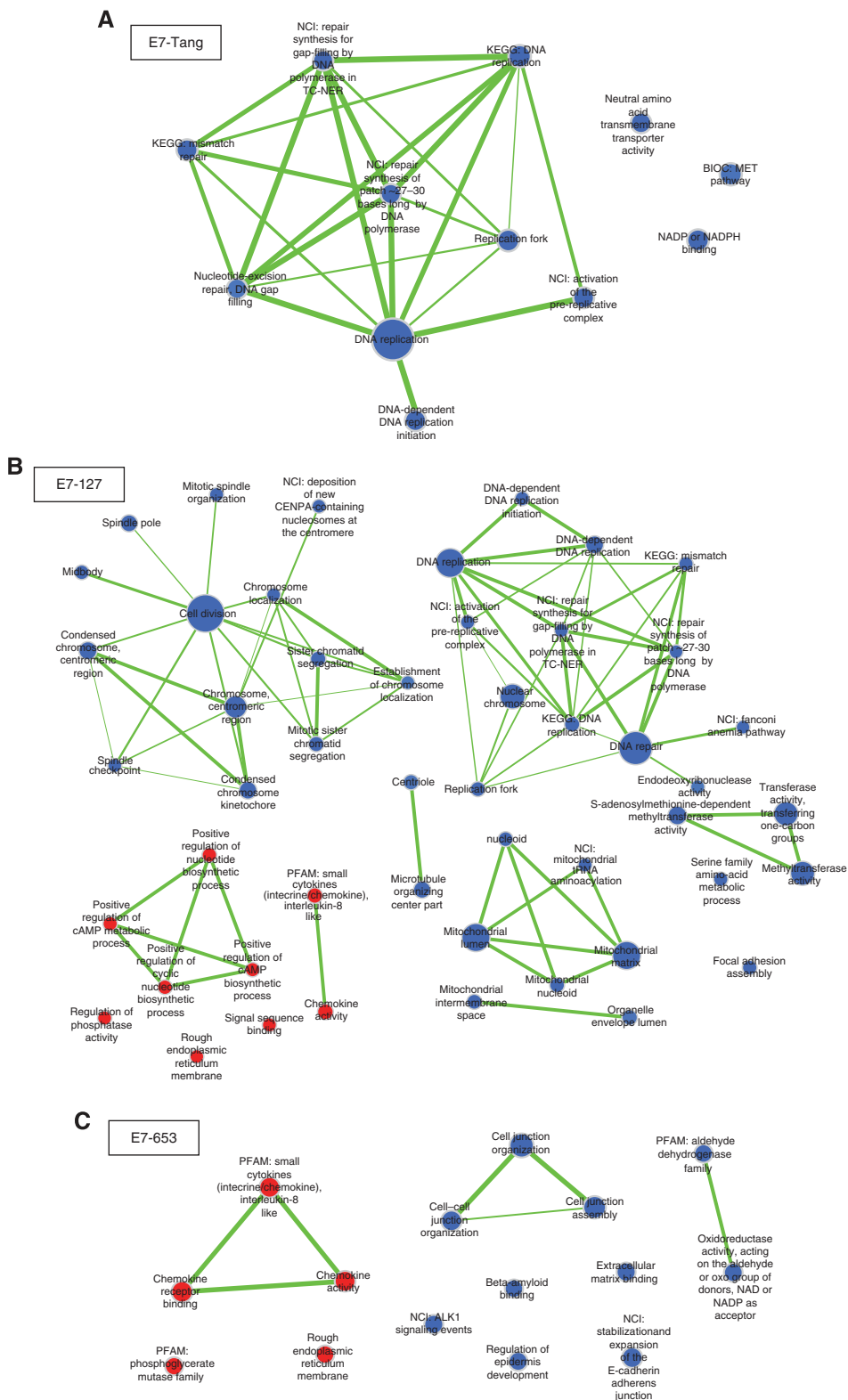
Figure 6. **Enriched gene sets by GSEA in NCx.** The maps show enriched gene sets for E7-Tang (**A**), E7–127 (**B**) and E7–653 (**C**) in NCx. Gene sets in upregulated genes are in red, whereas those in downregulated genes are in blue. Node size represents the gene-set size and line thickness represents the degree of overlap between gene sets.

However, intracellular abundance would not explain the effects of E7–653, which had as many predicted OTEs as E7–141 and E7-Tang. E7–653 may instead have shown greater sequence-dependent binding of innate immune receptors (Sioud, 2010).

In addition, our data suggested that there were indirect (downstream) OTEs that resulted from the other types of OTE. Such OTEs affected important cancer-associated pathways, with the enrichment for cell cycle regulation genes among

the mRNAs that were significantly downregulated following E7–127 treatment. These effects were likely to be downstream of E7–127-induced stimulation of innate immune pathways, as there was no enrichment for cell cycle pathway genes in the significantly downregulated SCR-containing mRNAs.

For E7–653, the indirect effects were cell-specific, being observed in NCx but not HaCaT. These OTEs were because of the sequences contained within the 8-mer 5′-CAGCAATA-3′, which were enriched in the 3′ UTRs of mRNAs downregulated by E7–653 in NCx. Interestingly, the 8-mer sequence was previously detected as a conserved regulatory motif in mammalian genomes (Xie *et al*, 2005). It does not correspond to any known miRNA binding site, but is thought to be an AU-rich element (ARE) motif (Xie *et al*, 2005). AREs are important in controlling gene expression through mRNA stability and degradation (Chen and Shyu, 1995) and can contribute to immune system control networks (Chen and Shyu, 1995; Schott and Stoecklin, 2010). E7–653 transfection of NCx may therefore cause mRNA degradation via this ARE motif. We found no evidence of upregulation in NCx or HaCaT of RNA-binding proteins, such as HuR, that are implicated in ARE-mediated effects of siRNAs and miRNAs (Schott and Stoecklin, 2010) (data not shown). Whether these observations truly represent ARE-mediated mRNA degradation, and the basis of the observed cell specificity of the E7–653 effects, therefore remain uncertain.

An interesting general observation from our study was that for all siRNAs tested there was relatively little overlap between the actual transcripts that were differentially expressed in NCx *vs* HaCaT cells. Nevertheless, all the significant siRNA-seed-dependent and immune-related OTEs that we observed were seen in both NCx and HaCaT. These results imply that although individual siRNAs do not consistently affect specific mRNA targets, there are still common effects on key cellular pathway(s). This may be partly due to downregulation of genes at different points in the pathway(s), potentially reflecting prevailing gene expression levels in the different cell types. On the other hand, significant ARE-mediated effects on transcript levels were only seen in NCx and not HaCaT. The overall differences between the two cell types may reflect their anatomical site of origin (cervix *vs* skin) and/or the immortalised phenotype of HaCaT. In either case, our data imply that it is advisable to use the most suitable cell type for *in vitro* investigations of OTEs of therapeutic siRNAs. When evaluating OTEs of siRNAs designed to treat cervical carcinoma, we consider the optimal cells to be HPV-negative NCx. The absence of the true siRNA target in such cells precludes any indirect effects of on-target gene depletion, which may be extensive and confound the identification of OTEs that are significant and clinically relevant in target-negative cells (Anderson *et al*, 2008).

We conclude that algorithms for predicting OTEs of therapeutic siRNAs on the basis of the numbers of SCR-containing mRNA targets are of limited value, as they do not predict actual seed-dependent OTEs accurately. Moreover, the number of actual OTEs observed does not dictate the significance of OTEs at a pathway level and does not predict the phenotypic effects. Our data indicate that the therapeutic siRNAs with high numbers of predicted OTEs should not necessarily be dismissed, and may even generate fewer actual seed-dependent and/or seed-independent OTEs than those with low numbers of predicted OTEs. This is exemplified by E7–141, which combined high potency with no significant OTEs *in vitro* and is a promising candidate for therapeutic targeting in cervical neoplasia. We suggest that, when designing siRNAs for therapeutic use, the potency of the siRNAs should be tested irrespective of the numbers of predicted OTEs. The most potent siRNAs should then undergo assessment of OTEs *in vitro*, using microarray and phenotypic analyses of the most relevant primary cell type.

## REFERENCES

Anderson EM, Birmingham A, Baskerville S, Reynolds A, Maksimova E, Leake D, Fedorov Y, Karpilow J, Khvorova A (2008) Experimental validation of the importance of seed complement frequency to siRNA specificity. *RNA* **14**(5): 853–861.

Arvey A, Larsson E, Sander C, Leslie CS, Marks DS (2010) Target mRNA abundance dilutes microRNA and siRNA activity. *Mol Syst Biol* **6**: 363.

Barbosa-Morais NL, Dunning MJ, Samarajiwa SA, Darot JF, Ritchie ME, Lynch AG, Tavare S (2010) A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data. *Nucleic Acids Res* **38**(3): e17.

Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* **136**(2): 215–233.

Bartonicek N, Enright AJ (2010) SylArray: a web server for automated detection of miRNA effects from expression data. *Bioinformatics* **26**(22): 2900–2901.

Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I (2001) Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* **125**(1-2): 279–284.

Birmingham A, Anderson E, Sullivan K, Reynolds A, Boese Q, Leake D, Karpilow J, Khvorova A (2007) A protocol for designing siRNAs with high functionality and specificity. *Nat Protoc* **2**(9): 2068–2078.

Boukamp P, Petrussevska RT, Breitkreutz D, Hornung J, Markham A, Fusenig NE (1988) Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *J Cell Biol* **106**(3): 761–771.

Chen CY, Shyu AB (1995) AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem Sci* **20**(11): 465–470.

Dunning MJ, Barbosa-Morais NL, Lynch AG, Tavare S, Ritchie ME (2008) Statistical issues in the analysis of Illumina data. *BMC Bioinformatics* **9**: 85.

Dunning MJ, Smith ML, Ritchie ME, Tavare S (2007) beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics* **23**(16): 2183–2184.

Falcon S, Gentleman R (2007) Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**(2): 257–258.

Fedorov Y, Anderson EM, Birmingham A, Reynolds A, Karpilow J, Robinson K, Leake D, Marshall WS, Khvorova A (2006) Off-target effects by siRNA can induce toxic phenotype. *RNA* **12**(7): 1188–1196.

Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**(1): 91–105.

Hammell M, Long D, Zhang L, Lee A, Carmack CS, Han M, Ding Y, Ambros V (2008) mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nat Methods* **5**(9): 813–819.

Herdman MT, Pett MR, Roberts I, Alazawi WO, Teschendorff AE, Zhang XY, Stanley MA, Coleman N (2006) Interferon-beta treatment of cervical keratinocytes naturally infected with human papillomavirus 16 episomes promotes rapid reduction in episome numbers and emergence of latent integrants. *Carcinogenesis* **27**(11): 2341–2353.

Jackson AL, Burchard J, Leake D, Reynolds A, Schelter J, Guo J, Johnson JM, Lim L, Karpilow J, Nichols K, Marshall W, Khvorova A, Linsley PS (2006) Position-specific chemical modification of siRNAs reduces 'off-target' transcript silencing. *RNA* **12**(7): 1197–1205.

Jackson AL, Linsley PS (2010) Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nat Rev Drug Discov* **9**(1): 57–67.

Jiang M, Milner J (2002) Selective silencing of viral gene expression in HPV-positive human cervical carcinoma cells treated with siRNA, a primer of RNA interference. *Oncogene* **21**(39): 6041–6048.

Jiang Y, Yin H, Zheng XL (2010) MicroRNA-1 inhibits myocardin-induced contractility of human vascular smooth muscle cells. *J Cell Physiol* **225**(2): 506–511.

Judge A, MacLachlan I (2008) Overcoming the innate immune response to small interfering RNA. *Hum Gene Ther* **19**(2): 111–124.

Kittler R, Surendranath V, Heninger AK, Slabicki M, Theis M, Putz G, Franke K, Caldarelli A, Grabner H, Kozak K, Wagner J, Rees E, Korn B, Frenzel C, Sachse C, Sonnichsen B, Guo J, Schelter J, Burchard J, Linsley PS, Jackson AL, Habermann B, Buchholz F (2007) Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies. *Nat Methods* **4**(4): 337–344.

Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**(1): 15–20.

Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**(7027): 769–773.

Magaldi TG, Almstead LL, Bellone S, Prevatt EG, Santin AD, DiMaio D (2012) Primary human cervical carcinoma cells require human papillomavirus E6 and E7 expression for ongoing proliferation. *Virology* **422**(1): 114–124.

Merico D, Isserlin R, Stueker O, Emili A, Bader GD (2010) Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* **5**(11): e13984.

Muralidhar B, Goldstein LD, Ng G, Winder DM, Palmer RD, Gooding EL, Barbosa-Morais NL, Mukherjee G, Thorne NP, Roberts I, Pett MR, Coleman N (2007) Global microRNA profiles in cervical squamous cell carcinoma depend on Drosha expression levels. *J Pathol* **212**(4): 368–377.

Okamura K, Lai EC (2008) Endogenous small interfering RNAs in animals. *Nat Rev Mol Cell Biol* **9**(9): 673–678.

Olejniczak M, Polak K, Galka-Marciniak P, Krzyzosiak WJ (2011) Recent advances in understanding of the immunological off-target effects of siRNA. *Curr Gene Ther* **11**(6): 532–543.

Palmer RD, Murray MJ, Saini HK, van Dongen S, Abreu-Goodger C, Muralidhar B, Pett MR, Thornton CM, Nicholson JC, Enright AJ, Coleman N (2010) Malignant germ cell tumors display common microRNA profiles resulting in global changes in expression of messenger RNA targets. *Cancer Res* **70**(7): 2911–2923.

Pattillo RA, Hussa RO, Story MT, Ruckert AC, Shalaby MR, Mattingly RF (1977) Tumor antigen and human chorionic gonadotropin in CaSki cells: a new epidermoid cervical cancer cell line. *Science* **196**(4297): 1456–1458.

Pett MR, Alazawi WO, Roberts I, Dowen S, Smith DI, Stanley MA, Coleman N (2004) Acquisition of high-level chromosomal instability is associated with integration of human papillomavirus type 16 in cervical keratinocytes. *Cancer Res* **64**(4): 1359–1368.

Phalon C, Rao DD, Nemunaitis J (2010) Potential use of RNA interference in cancer therapy. *Expert Rev Mol Med* **12**: e26.

Schott J, Stoecklin G (2010) Networks controlling mRNA decay in the immune system. *Wiley Interdiscip Rev RNA* **1**(3): 432–456.

Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**(7209): 58–63.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**(11): 2498–2504.

Sioud M (2010) Recent advances in small interfering RNA sensing by the immune system. *N Biotechnol* **27**(3): 236–242.

Smotkin D, Wettstein FO (1986) Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein. *Proc Natl Acad Sci USA* **83**(13): 4680–4684.

Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**: Article3.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**(43): 15545–15550.

Tang S, Tao M, McCoy Jr JP, Zheng ZM (2006) Short-term induction and long-term suppression of HPV16 oncogene silencing by RNA interference in cervical cancer cells. *Oncogene* **25**(14): 2094–2104.

Ui-Tei K, Naito Y, Zenno S, Nishi K, Yamato K, Takahashi F, Juni A, Saigo K (2008) Functional dissection of siRNA sequence by systematic DNA substitution: modified siRNA with a DNA seed arm is a powerful tool for mammalian gene silencing with significantly reduced off-target effect. *Nucleic Acids Res* **36**(7): 2136–2151.

Vaishnaw AK, Gollob J, Gamba-Vitalo C, Hutabarat R, Sah D, Meyers R, de Fougerolles T, Maraganore J (2010) A status report on RNAi therapeutics. *Silence* **1**(1): 14.

van Dongen S, Abreu-Goodger C, Enright AJ (2008) Detecting microRNA binding and siRNA off-target effects from expression data. *Nat Methods* **5**(12): 1023–1025.

Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, Snijders PJ, Peto J, Meijer CJ, Munoz N (1999) Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol* **189**(1): 12–19.

Winder DM, Chattopadhyay A, Muralidhar B, Bauer J, English WR, Zhang X, Karagavriilidou K, Roberts I, Pett MR, Murphy G, Coleman N (2011) Overexpression of the oncostatin M receptor in cervical squamous cell carcinoma cells is associated with a pro-angiogenic phenotype and increased cell motility and invasiveness. *J Pathol* **225**(3): 448–462.

Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* **434**(7031): 338–345.

Yamato K, Egawa N, Endo S, Ui-Tei K, Yamada T, Saigo K, Hyodo I, Kiyono T, Nakagawa I (2011) Enhanced specificity of HPV16 E6E7 siRNA by RNA-DNA chimera modification. *Cancer Gene Ther* **18**(8): 587–597.

Yamato K, Yamada T, Kizaki M, Ui-Tei K, Natori Y, Fujino M, Nishihara T, Ikeda Y, Nasu Y, Saigo K, Yoshinouchi M (2008) New highly potent and specific E6 and E7 siRNAs for treatment of HPV16 positive cervical cancer. *Cancer Gene Ther* **15**(3): 140–153.

Supplementary Information accompanies this paper on British Journal of Cancer website (http://www.nature.com/bjc)