FACULTY OF HEALTH AND MEDICAL SCIENCES
UNIVERSITY OF COPENHAGEN
COPENHAGEN CENTER FOR GLYCOMICS
DEPARTMENT OF ODONTOLOGY

Fast and Sensitive Detection of Indels Induced by Precise Gene Targeting

Eric P. Bennett[1], Zhang Yang [1,2], Catharina Steentoft[1], Camilla Hauge[1,2,3], Lars Hansen[1,2], Allan Lind Thomsen[1], Francesco Nicola[3], Morten Frödin[3], Malene B. Vester-Christensen[1], Henrik Clausen[1,2] and Hans H. Wandall[1]

[1] Copenhagen Center for Glycomics, Departments of Cellular and Molecular Medicine and School of Dentistry, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, 2200 Copenhagen N, Denmark
[2] Novo Nordisk Foundation Center for Biosustainability, Danish Technical University, Lyngby, Denmark
[3] Biotech Research and Innovation Centre, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen N, Denmark

Conclusion

The indel detection by amplicon analysis strategy (coined **IDAA^{TM}**) presented here enables sensitive, precise and reliable identification of indels in a high throughput mode providing detailed information of cutting efficiency, size and nature of allelic variants generated by any of the available precise gene editing technologies (**Fig.1**). The IDAA^{TM} strategy is user friendly and easily implemented in any standard laboratory, and we expect that this screening tool will greatly advance implementation and use of precise gene targeting.

Results

Current approaches available for identification of indels include: i) enzyme mismatch cleavage (EMC) assays[1], which do not provide sensitive, reliable and accurate identification of the induced indel's (**Fig.2**); and ii) Sanger or next generation DNA sequencing, which is costly, time and labor intensive, and poorly suited for high throughput screening of hundreds or thousands of clones often required to select for desirable multi-allelic editing events that often occur at low frequency (**Fig. 4**).

Here, we report a novel strategy that combines use of a simple amplicon labelling strategy (**Fig.1b**) with the high throughput capability of DNA fragment analysis by automated Capillary Electrophoresis[2] (**Fig.1c**) for simple detection and characterization of indels induced by precise gene targeting. The strategy is coined IDAA^{TM} for Indel Detection by Amplicon Analysis, and we demonstrate that IDAA is suitable for detecting indels in both cell pools with low efficiency targeting and single sorted cells. Furthermore, we show that IDAA^{TM} is ideally suited for high throughput detection of indels down single base events (**Fig.2-7**), estimation of "cutting" efficiencies of targeting tools (**Fig.3**), and evaluation of off-target events at candidate loci (**Fig.6-7**).

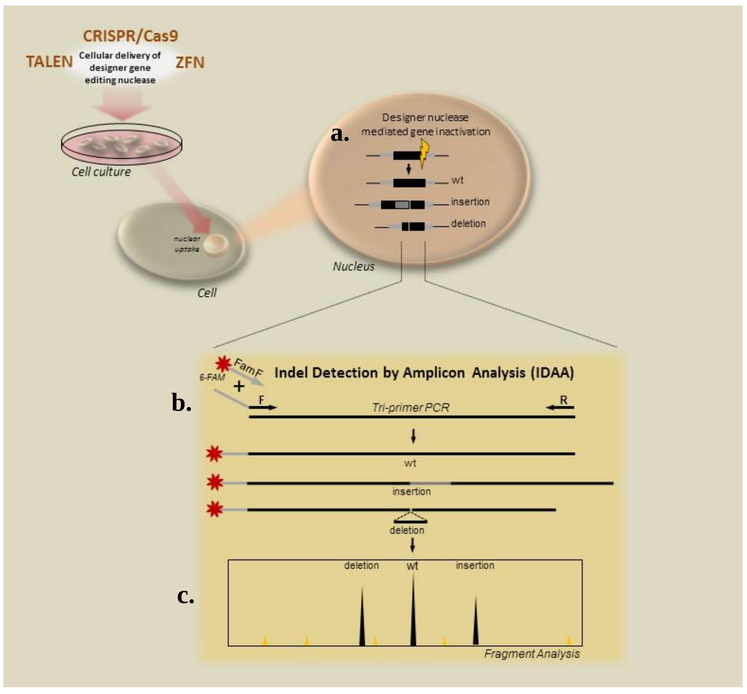

Figure 1. Schematic depiction of the IDAA^{TM} strategy*. **Panel a:** Precise gene targeting creates double-stranded breaks that through NHEJ introduce indels at the target site. **Panel b:** Tri-primer PCR^ of the target region accomplished by use of target specific primers (F/R) flanking the target site and a universal 5'-FAM labelled primer (FamF) specific for a 5'-overhang sequence attached to primer F. Tri-primer PCR results in FAM amplicon labelling. **Panel c:** Fluorescently labelled amplicons containing the indels are detected by fragment analysis[2]. Axis represent fluorescence intensity (FI) and amplicon size in base pairs.

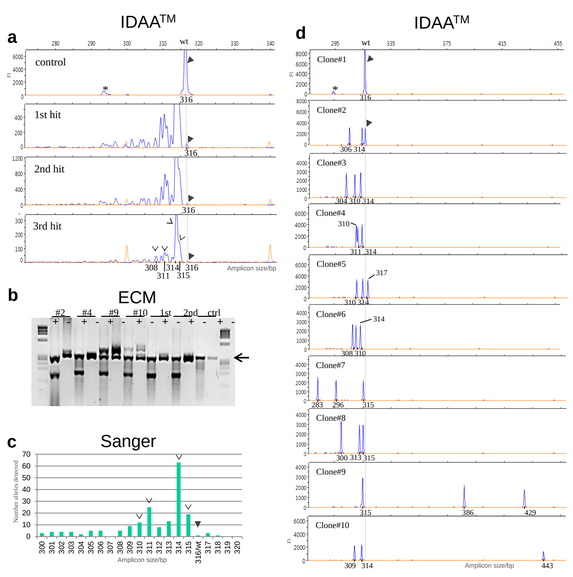

Figure 5. Tri-allelic K562 *KRAS* CRISPR/Cas9 targeting evaluated by IDAA^{TM}: Tri-allelic K562 *KRAS* gene was targeted using a plamsid expressing Cas9-2A-GFP and *KRAS* gRNA as previously described[4] and indels were detected after 3 consecutive rounds *KRAS* CRISPR/Cas9 targeting followed by FACS sorting of GFP positive cells, cell expansion and re-transfection (1st, 2nd and 3rd hit of cells) like previously described[4]. **Panel a:** IDAA^{TM} result of 1st, 2nd and 3rd rounds of transfected cells day 2 post transfection. Untransfected cell control is shown in upper panel. The appearance of an unspecific minor peak observed in this assay is indicated by an asterix. The position of the wt peak is indicated by filled arrow heads. Notably the wt peak is significantly diminished after the 3rd hit. Dominant peaks are marked by open arrow heads. Amplicon peaks are shown in blue and the LIZ500 standard in orange. **Panel b:** EMC/T7-assay results of representative 3rd hit single cell sorted clones and 1st and 2nd hit cell pools. Phi-X marker is positioned in the flanking lanes. Arrow indicates the major uncleaved amplicon detected. **Panel c:** Sanger results from 96 single sorted clones after 3rd hit. All Sanger detected indels from 96 clones are summarized. Note the comparable profile for the indels detected in the pool of cells shown in **panel a** marked by open arrow heads. **Panel d**: Representative IDAA^{TM} results from 3rd hit single cell clones displayed in **panel b**. Notable only one wt allel was detected.

Supplementary Table II "Off target" mismatch distribution at top 24 sites

	*	strand	gRNA gene target´s	
wt :	GAATATGTGAGTGTGGATGGAGG	: 23	−	Cosmc
1 :	GAATATGTGAGTGTGGAGGAGG	: 23	−	ras-related protein Rab-33A-like
2 :	GAATATGTGTGTGGGAGGAGG	: 23	−	intraflagellar transport protein 81
3 :	GAGTGTGTGTGTGGATGGAGG	: 23	−	unplaced genomic scaffold 1437
4 :	GAATATGTGAGTGTGAGTGGAGG	: 23	−	unplaced genomic scaffold 1699
5 :	GAATATTGTGTGTGGATGGAGG	: 23	+	unplaced genomic scaffold 300
6 :	GAAGGTGTGAGTGTGGATGGAGG	: 23	−	unplaced genomic scaffold 1586
7 :	GAATATGTGAGTGTGCATGGAGG	: 23	+	unplaced genomic scaffold 867
8 :	GAATGTGTGAGTGTGGTGGAGG	: 23	−	unplaced genomic scaffold 3676
9 :	GAAGAGGTGAGTGTGGATGGAGG	: 23	+	unplaced genomic scaffold 1413
10 :	GAGGTGTGAGTGTGCATGGAGG	: 23	+	unplaced genomic scaffold 6488
11 :	GAAGGTGTGAGTGTGGATGGAGG	: 23	−	unplaced genomic scaffold 3651
12 :	GAAGATGTGTGTGTGATGGTGG	: 23	−	hypothetical protein LOC100752970
13 :	GAGGAGGTGAGTGTGGTGGAGG	: 23	+	prospero homeobox protein 1
14 :	GAATGTGTGAATGTGTATGGAGG	: 23	+	cadherin-13
15 :	GAATATGTGTGTGTGTTGGAGG	: 23	−	polycystic kidney disease protein 1-L2
16 :	GAATGTGTGTGTGTGGATGGTGG	: 23	+	S-adenosylmethionine mitochondrial protein
17 :	GAATTTGTGAGTGTGGATGAGG	: 23	+	bone morphogenetic protein 2-like
18 :	GAATATGTGAGTGTTGGAGGAGG	: 23	+	interferon alpha/beta receptor 2-like
19 :	GAAGAGGTGAGTGTGCATGGAGG	: 23	+	disabled homolog 1-like
20 :	GAATGTGTGTGTGTGGAGGAGG	: 23	+	disabled homolog 1-like
21 :	GAGGTGTGTGTGTGTGGATGGAGG	: 23	−	E3 ubiquitin-protein ligase RNF216-like
22 :	GAATGTGTGTGTGTGCATGAGG	: 23	+	mitogen-activated protein kinase MLT-like
23 :	GAATATGTGCGTGTTGGATGGAGG	: 23	−	importin-11
24 :	GAATAATGTGTGTGTGGTGGAGG	: 23	−	S1 RNA-binding domain-containing protein

*indicates the last base in gRNA preceeding the PAM seed sequence

Figure 4. Schematic depiction of the labelled ZFN targeting vector for GALNT6 (Dual-*GALNT6*-ZFN). **Panel a:** GFP, 2A peptides (2A1 and 2A2), 3xFLAG (3xF), nuclear localization signal (NLS), ZFN1 and ZFN2 for *GALNT6* (GALNT6ZFN1 & 2), and 3xMyc (3xM) are indicated. **Panel b:** SDS-PAGE Western blot analysis of K562 cells harvested day 1 (D1), 2 (D2) and 5 (D5) after nucleofection with the Dual-*GALNT6*-ZFN plasmid targeting vector. Blots were reacted with anti-GFP, anti-FLAG, anti-Myc or anti-actin antibodies as indicated. Arrow heads indicate reactive bands with expected mobilities. For control transfections (C) K562 cells were transfected with monomeric Flag-tagged original ZFNs (Sigma-Aldrich) and harvested day 1. **Panel c:** IDAA^{TM} of a HepG2 cell pool at day 2 post nucleofection with Dual-*GALNT6*-ZFN. The low efficiency and prevalence of +/-1bp indels were confirmed by MiSeq deep sequencing (**panel d**). Indel percentages (shown in parenthesis) were calculated from peak areas of the summed indel peaks relative to the total peak area. **Panel d:** Number of indels detected by MiSeq with indel sizes ranging from -5bp to +5bp shown. Note the profile and indel frequencies for the -1bp, 0, +1 bp matches the IDAA^{TM} profiles shown **panel c**.

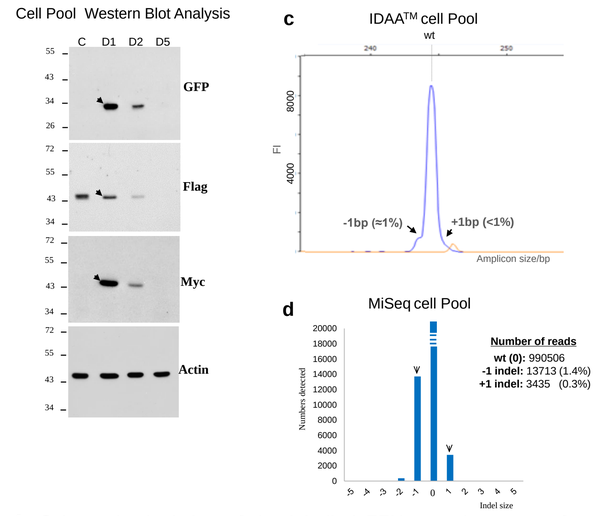

Figure 6: Panel a: Schematic depiction of the gRNA2 target region in the *Cosmc* gene locus and candidate off-target loci with 1-4 mismatches. **Panel b:** IDAA^{TM} of the top off-target candidate region with 2 mismatches in 10 single CHO cell clones. The *Cosmc* Cas9/gRNA2 targeted single cell sorted clones were from the experiment presented in **Figure 3**. Position of the intact amplicon indicated by 0 above peak (0bp indel). No off-target events were detected and this was confirmed by Sanger sequencing. LIZ600 marker positions are shown as unmarked peaks within diagrams.

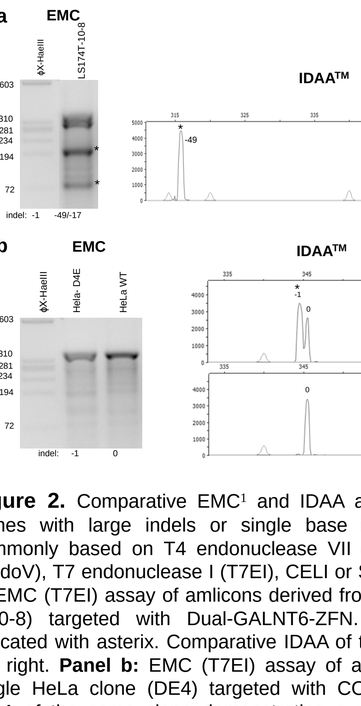

Figure 2. Comparative EMC[1] and IDAA analysis of ZFN targeted clones with large indels or single base indel. EMC assays are commonly based on T4 endonuclease VII (T4E7), endonuclease V (EndoV), T7 endonuclease I (T7EI), CELI or Surveyor nuclease. Panel **a:** EMC (T7EI) assay of amlicons derived from a single LS174T clone (#10-8) targeted with Dual-GALNT6-ZFN. Cleaved products are indicated with asterix. Comparative IDAA of the same clone shown to the right. **Panel b:** EMC (T7EI) assay of amplicons derived from a single HeLa clone (DE4) targeted with COSMC-ZFN. Comparative IDAA of the same clone demonstrating a monoallelic -1bp deletion (indicated with and asterix) shown to the left, relative to the intact HeLa WT peak(0). Unmarked minor light grey peaks represent the GSLIZ600 standard.

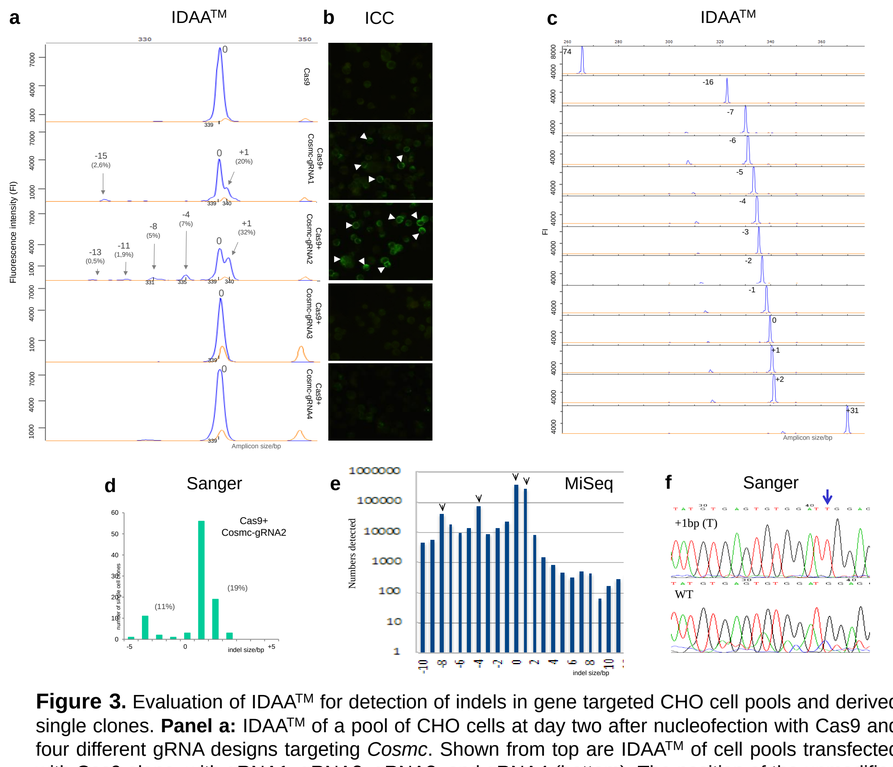

Figure 3. Evaluation of IDAA^{TM} for detection of indels in gene targeted CHO cell pools and derived single clones. **Panel a:** IDAA^{TM} of a pool of CHO cells at day two after nucleofection with Cas9 and four different gRNA designs targeting *Cosmc*. Shown from top are IDAA^{TM} of cell pools transfected with Cas9 alone, with gRNA1, gRNA2, gRNA3, and gRNA4 (bottom). The position of the unmodified wild type amplicon peak is indicated (0) and amplicon sizes (in bp) as determined by Peak Scanner software are indicated below peaks. Indel sizes determined (in bp) are shown above the most prominent peaks together with cutting efficiencies (calculated from peak area relative to total peak area) in percentage. Total cutting efficiency for gRNA1 and gRNA2 were estimated to 23% and 46%, respectively, while gRNA3 and gRNA4 were inactive. The relative frequency of indels produced by gRNA2 was confirmed by MiSeq deep sequencing (**panel e**) which further revealed that the predominant +1 insertion was an thymine insertion at a position three bases upstream of the PAM sequence (**panel f**). The GSLIZ500 standard peaks are shown in orange. **Panel b:** Comparative ICC analysis of the corresponding cell pools shown in (**a**) seven days after nucleofection with a monoclonal antibody (5F4) detecting the de novo induction of truncated O-glycans as a result of complete inactivation of *Cosmc*[3] (arrow heads indicate single positive cells in pool). **Panel c:** Representative IDAA^{TM} analysis of single cell clones showing the 1bp resolution power of IDAA^{TM}. F. **Panel d:** Analysis of indels by Sanger sequencing showing distribution in the -5bp to +5bp range of individual single cell clones. **Panel e:** Number of indels detected by MiSeq deep NGS, indel sizes ranging from -10bp to +10bp are shown. Note the profile for the -4bp, 0, +1 bp matches the IDAA^{TM} and Sanger profiles shown in **panels a** and **d**. Y-axis is logarithmic. **Panel f:** Single cell clone Sanger identification of the predominant +1bp insertion identified by IDAA^{TM,} as a thymine insertion at a position three bases upstream of the PAM sequence.

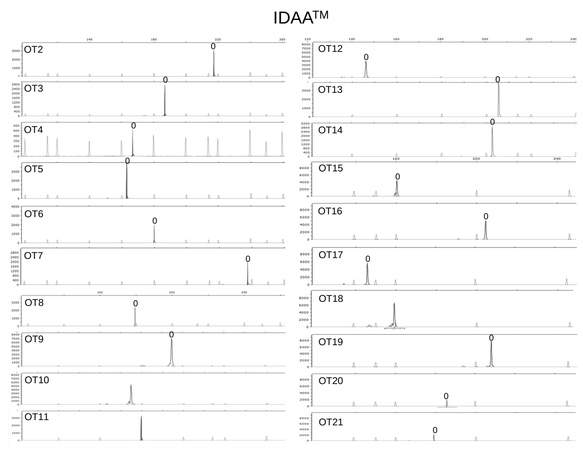

Figure 7. IDAA^{TM} of the top 21 off-target (OT) loci are shown for a single CHO clone (#7) targeted with *Cosmc* Cas9-gRNA2 from the experiment described for **Figure 3**. IDAA^{TM} of all 21 off-targets was performed on additional 9 CHO clones, and no off-target events were identified. IDAA^{TM} OT1 results are shown in **Figure 6**. Position of the only detected intact amplicon (representing the unmodified off target) is indicated by 0 above peak (0bp indel). Results were verified by Sanger sequencing of all amplicons analyzed. LIZ600 marker positions are shown as unmarked peaks within diagrams. Note, that the relative differences in amplicon product yields for the different targets shown, give rise to considerable variation in the intensities of the LIZ600 marker shown as unmarked peaks in the respective chromatograms.

1. Huang, M.C., Cheong, W.C., Lim, L.S. and Li, M.-H. (2012) A simple, high sensitivity mutation screening using Ampligase mediated T7 endonuclease I and Surveyor nuclease with microfluidic capillary electrophoresis. *Electrophoresis*, **33**, 788–96.
2. Andersen, P.S., Jespersgaard, C., Vuust, J., Christiansen, M. and Larsen, L.A. (2003) Capillary electrophoresis-based single strand DNA conformation analysis in high-throughput mutation screening. *Hum. Mutat.*, **21**, 455–65.
3. Steentoft, C., Vakhrushev S.Y., Vester-Christensen M.B., Schjoldager K.T., Kong Y, **Bennett E.P.**, Mandel U, Wandall H.H., Levery S.B., Clausen H. (2011) Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat. Methods* **8,** 977–82.
4. Duda, K., Lonowski, L. a, Kofoed-Nielsen, M., Ibarra, A., Delay, C.M., Kang, Q., Yang, Z., Pruett-Miller, S.M., **Bennett, E.P.**, Wandall, H.H., et al. (2014) High-efficiency genome editing via 2A-coupled co-expression of fluorescent proteins and zinc finger nucleases or CRISPR/Cas9 nickase pairs. *Nucleic Acids Res.*, 42(10):e84.
5. Yang, Z., Steentoft C., Hauge C., Hansen L., Lind-Thomse A., Niola F., Vester-Christensen M.B., Frodin M., Clausen H., Wandall H.H. and **Bennett E.P.** (2015) Fast and Sensitive Detection of Indels Induced by Precise Gene Targeting. *NAR*, doi=10.1093/nar/gkv126, *in press.*

^ All IDAA^{TM} primers are available from TAG Copenhagen A/S, Denmark (http://tagc.dk/) in a kit format (**IDAA^{TM}-kit**).

***** A provisional patent application covering parts of the technology has been filed.

Funding: This work was supported by The Danish Research Councils, The University of Copenhagen program of excellence 2016 [CDO2016], The Danish National Research Foundation [DNRF107], and the Novo Nordisk Foundation.