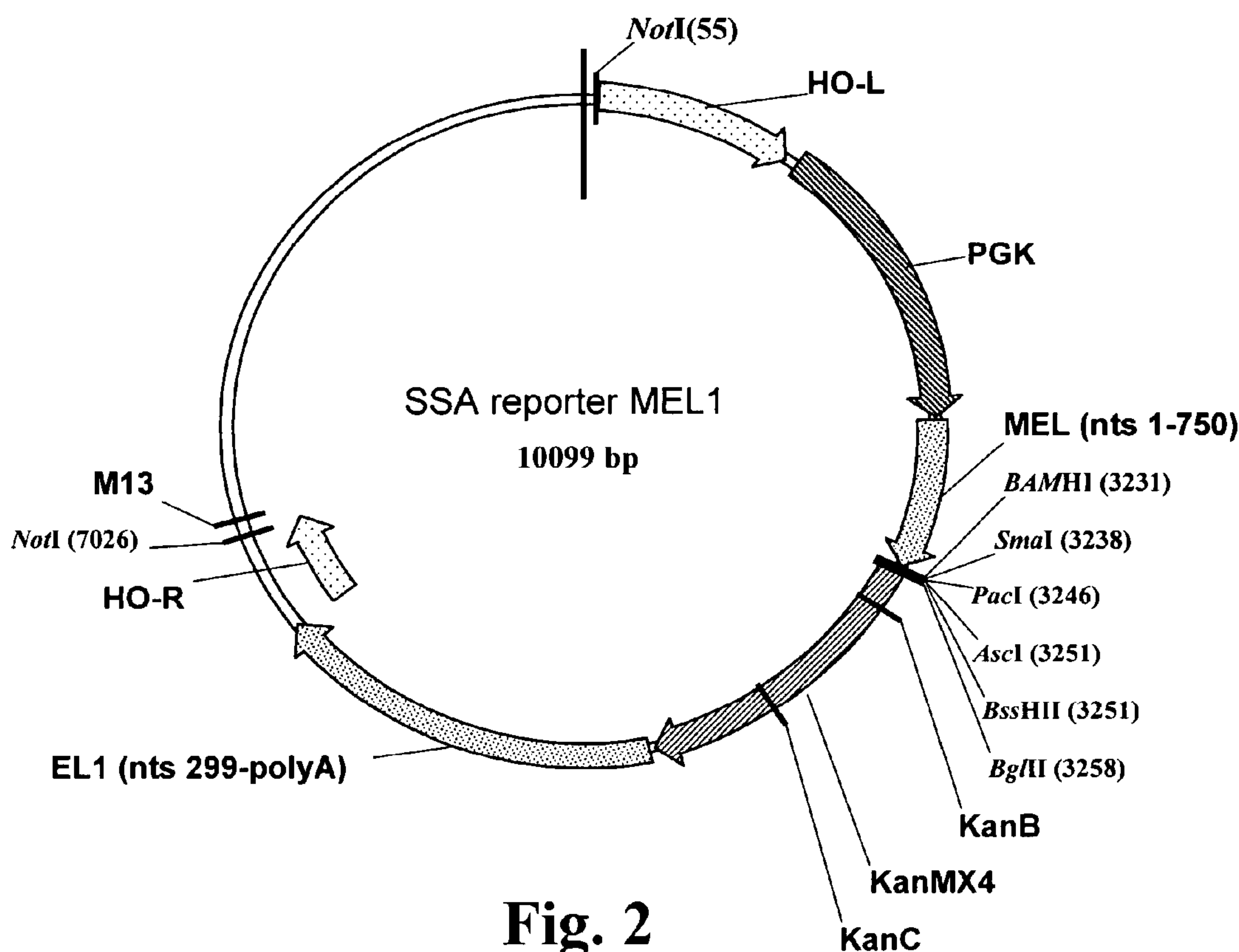




(86) Date de dépôt PCT/PCT Filing Date: 2008/09/25
(87) Date publication PCT/PCT Publication Date: 2009/04/02
(85) Entrée phase nationale/National Entry: 2010/03/19
(86) N° demande PCT/PCT Application No.: US 2008/011087
(87) N° publication PCT/PCT Publication No.: 2009/042163
(30) Priorité/Priority: 2007/09/27 (US60/995,566)

(51) Cl.Int./Int.Cl. *C12Q 1/34* (2006.01),
G01N 33/573 (2006.01)
(71) Demandeur/Applicant:
SANGAMO BIOSCIENCES, INC., US
(72) Inventeurs/Inventors:
DOYON, YANNICK, US;
URNOV, FYODOR, US
(74) Agent: KIRBY EADES GALE BAKER

(54) Titre : IDENTIFICATION IN VIVO RAPIDE DE NUCLEASES BIOLOGIQUEMENT ACTIVES
(54) Title: RAPID IN VIVO IDENTIFICATION OF BIOLOGICALLY ACTIVE NUCLEASES



(57) Abrégé/Abstract:

Disclosed herein are methods and compositions for rapidly identifying and ranking nucleases for specific cleavage of a target sequence.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
2 April 2009 (02.04.2009)

PCT

(10) International Publication Number
WO 2009/042163 A2

(51) International Patent Classification:
C12Q 1/34 (2006.01) *G01N 33/573* (2006.01)

(74) Agents: **PASTERNAK, Dahna S.** et al.; Robins & Pasternak LLP, 1731 Embarcadero Rd., Suite 230, Palo Alto, CA 94303 (US).

(21) International Application Number:
PCT/US2008/011087

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(22) International Filing Date:
25 September 2008 (25.09.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/995,566 27 September 2007 (27.09.2007) US

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): **SANG-AMO BIOSCIENCES, INC.** [US/US]; Point Richmond Tech Center, 501 Canal Blvd., Suite A100, Richmond, CA 94804 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **DOYON, Yannick** [CA/US]; 6501 Barrett Ave., El Cerrito, CA 94530 (US). **URNOV, Fyodor** [US/US]; 1400 Pinnacle Court #313, Point Richmond, CA 94801 (US).

Published:

— without international search report and to be republished upon receipt of that report

(54) Title: RAPID IN VIVO IDENTIFICATION OF BIOLOGICALLY ACTIVE NUCLEASES

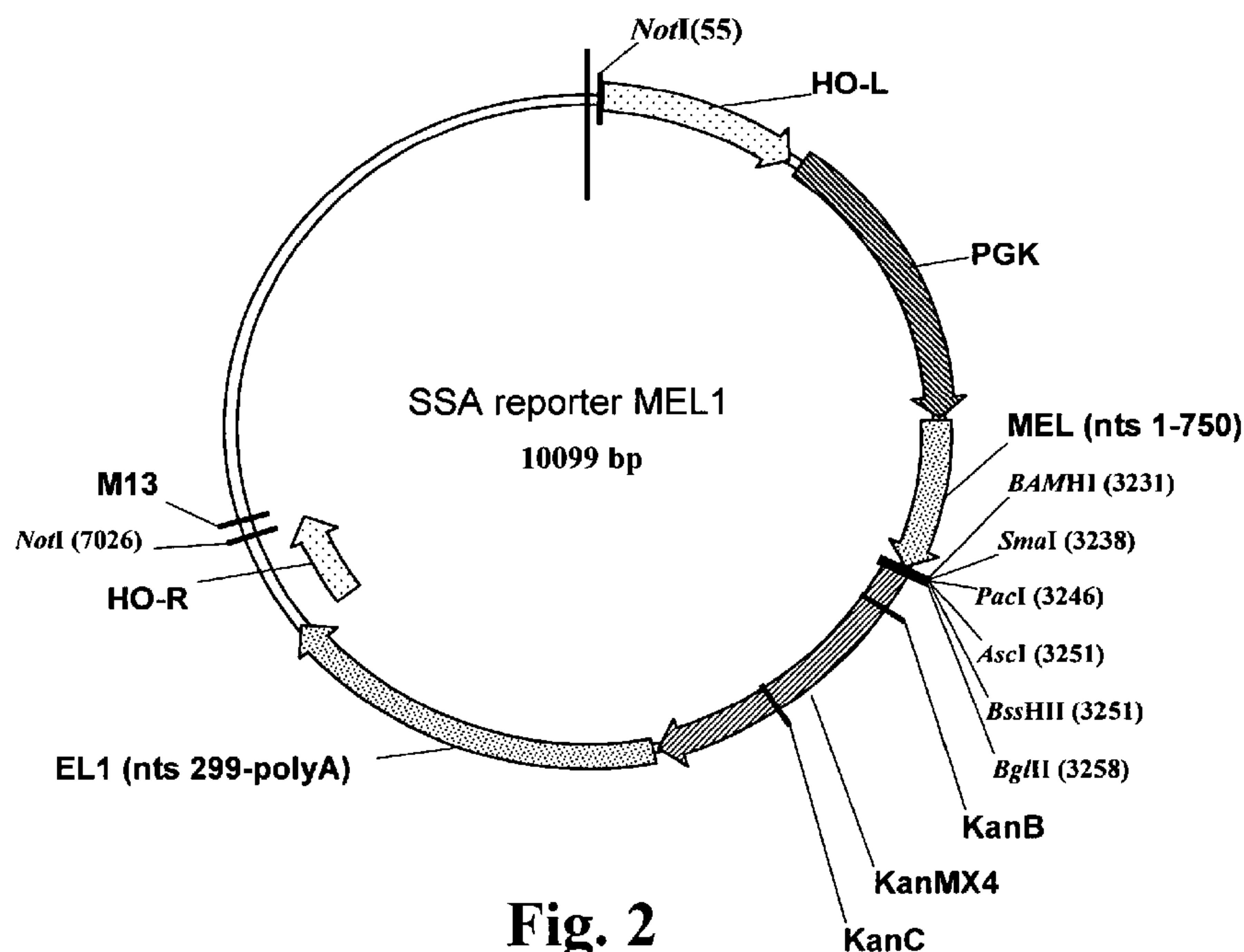


Fig. 2

(57) Abstract: Disclosed herein are methods and compositions for rapidly identifying and ranking nucleases for specific cleavage of a target sequence.

WO 2009/042163 A2

**RAPID *IN VIVO* IDENTIFICATION OF
BIOLOGICALLY ACTIVE NUCLEASES**

5

STATEMENT OF RIGHTS TO INVENTIONS
MADE UNDER FEDERALLY SPONSORED RESEARCH

[0001] Not applicable.

TECHNICAL FIELD

10

[0002] The present disclosure is in the fields of genome engineering and nuclease identification.

BACKGROUND

15

[0003] Nucleases, including zinc finger nucleases and homing endonucleases such as SclI, that are engineered to specifically bind to target sites have been shown to be useful in genome engineering. For example, zinc finger nucleases (ZFNs) are proteins comprising engineered site-specific zinc fingers fused to a nuclease domain. Such ZFNs have been successfully used for genome modification in a variety of different species. See, for example, United States Patent Publications 20030232410; 20050208489; 20050026157; 20050064474; 20060188987; 20060063231; and International Publication WO 07/014275, the disclosures of which are incorporated by reference in their entireties for all purposes. These ZFNs can be used to create a double-strand break (DSB) in a target nucleotide sequence, which increases the frequency of homologous recombination at the targeted locus more than 1000-fold.

25

In addition, the inaccurate repair of a site-specific DSB by non-homologous end joining (NHEJ) can also result in gene disruption. Creation of two such DSBs results in deletion of arbitrarily large regions.

30

[0004] Currently, ZFNs specific for particular targets are generally identified using *in vitro* assays used to identify engineered zinc finger proteins. See, e.g., U.S. Patent Publication No. 20050064474. However, these *in vitro* assays are time and labor intensive. Furthermore, although *in vitro* methods accurately identify ZFPs with the desired binding activity, the architecture of ZFNs and the chromatin infrastructure over the target locus in living cells may in some instances hinder the capacity of these *in vitro* assays to accurately predict *in vivo* ZFN activity.

[0005] *In vivo* screening assays, particularly in yeast host cells, have been used to select homing endonucleases that bind to target sites other than their cognate binding site. See, e.g., Chames et al. (2005) *Nucleic Acids Res* 33(20):e178; Arnould et al. (2006) *J. Mol. Biol.* 355:443-458; and U.S. Patent Publication Nos.

5 20070117128; 20060206949; 20060153826; 20060078552; and 20040002092.

However, such methods have not been broadly applied to any nuclease, including zinc finger nucleases. Moreover, previously described *in vivo* methods do not identify biologically active nucleases from a panel of nucleases known to bind to a specific target site, nor from a panel of nucleases known to bind to a set of sites within a particular genomic region. Rather, these previously-described *in vivo* screening assays utilize a randomly generated library of mutant homing endonucleases to identify proteins which bind to a particular, specific target site. Thus, previously-described assays do not predict *in vivo* functionality from a collection of nucleases known to bind to a particular target, nor from a collection of nucleases known to bind to a set of distinct targets within a broader genomic region. Nor do these assays accurately determine which nucleases are least toxic to the host cell.

[0006] Thus, there remains a need for additional assays to identify specific nucleases, particularly high throughput *in vivo* assays that identify functional, specifically-targeted nucleases.

20

SUMMARY

[0007] The present disclosure relates to development of nucleases, for example engineered meganucleases and zinc finger nuclease (ZFNs). Specifically, described herein are compositions and methods for the efficient screening, identification, and ranking of biologically active engineered nucleases. In addition, the assay systems described herein also allow for rapid toxicity screening of such nucleases.

[0008] The rapid identification of highly active and specific lead nucleases for a particular target gene as described herein significantly alleviates the obstacles associated with repetitive and time-consuming experiments typically performed in diverse cell types and organisms.

[0009] In one aspect, described herein is a reporter construct for detecting double-stranded cleavage of a target sequence by one or more nucleases. The reporter construct comprises overlapping and non-functional sequences of a reporter

gene separated by a target sequence recognized by the nuclease. The 5' region of the reporter gene may be operably linked to a constitutive or inducible promoter. The reporter gene may encode an enzymatic protein, for example Mel1. Expression of the reporter construct in a host cell results in a signal that is measurable by suitable

5 assays, for example by colorimetric or enzymatic assays performed on intact or lysed cells. In certain embodiments, activity of the reporter gene is determined by assaying levels of a secreted protein (*e.g.*, the product of the reporter gene itself or a product produced directly or indirectly by an active reporter gene product). In certain
10 embodiments, the reporter construct also comprises regions of homology flanking the discontinuous reporter gene sequences and/or a selectable marker. The regions of homology may be to any region of a host cell genome, for example the HO locus in yeast. Optionally, a second reporter gene is also included, for example a reporter that is transcribed only in the presence of double-stranded breaks. In certain
15 embodiments, the reporter construct comprises a construct as shown in Fig. 2 or Fig. 9.

[0010] In another aspect, described herein is a host cell (or population of host cells) comprising any of the reporter constructs described herein. The host cell typically includes the cellular machinery (endogenous or exogenous) for processing a double-stranded break to create overlapping single-stranded sequences that are
20 repaired via single-stranded annealing repair. In certain embodiments, the host cell is a yeast cell, for example *S. cerevisiae*. The reporter construct may be transiently expressed in the host cell. Alternatively, the reporter construct is stably integrated into the genome of the host cell.

[0011] In yet another aspect, methods of identifying a nuclease that induce(s)
25 cleavage at a specific target site are provided. In certain embodiments, the methods comprise introducing one or more nuclease and/or one or more nuclease-expression constructs encoding a nuclease or a pair of nucleases into a host cell comprising a reporter construct as described herein, the reporter construct comprising a target
30 sequence recognized by the nuclease(s); incubating the cells under conditions such that the nuclease(s) are expressed; and measuring the levels of reporter gene expression in the cells, wherein increased levels of reporter gene expression are correlated with increased nuclease-induced cleavage of the target sequence. The nuclease may comprise, for example, a non-naturally occurring DNA-binding domain (*e.g.*, an engineered zinc finger protein or an engineered DNA-binding domain from a

homing endonuclease). In certain embodiments, the nuclease is a zinc finger nuclease (ZFN) or pair of ZFNs.

[0012] In yet another aspect, methods of ranking a panel of nucleases for their cleavage-inducing activity at a specific target site are provided. The methods

5 comprise introducing a nuclease of the panel and/or expression constructs encoding nuclease of the panel into separate host cells, the host cells each comprising a reporter construct as described herein, the reporter construct comprising a target sequence recognized by the nuclease(s); incubating the cells under conditions such that the nuclease(s) are expressed; measuring the levels of reporter gene expression in the
10 cells; and ranking the nuclease(s) according to levels of reporter gene activity induced in the host cell. In certain embodiments, the nuclease comprises a ZFN or ZFN pair. In other embodiments, the nuclease comprises a homing endonuclease with an engineered DNA-binding domain and/or a fusion of a DNA-binding domain of a homing nuclease and a cleavage domain of a heterologous nuclease.

15 [0013] In another aspect, methods of predicting the *in vivo* cleavage activity of a nuclease are provided. The methods comprise introducing the nuclease and/or expression constructs encoding a nuclease into a host cell comprising a reporter construct as described herein, the reporter construct comprising a target sequence recognized by the nuclease; incubating the cells under conditions such that the
20 nuclease is expressed; and measuring the levels of reporter gene expression in the cells; wherein higher levels of reporter gene expression are predictive of a nuclease that will be active *in vivo*. In certain embodiments, the nuclease comprises a ZFN or ZFN pair. In other embodiments, the nuclease comprises a homing endonuclease with an engineered DNA-binding domain and/or a fusion of a DNA-binding domain of a
25 homing nuclease and a cleavage domain of a heterologous nuclease.

[0014] In yet another aspect, methods of determining toxic effects on a host cell caused by a nuclease are provided. The methods comprise introducing a nuclease and/or one or more expression construct(s) encoding one or more nucleases into a host cell; incubating the cells under conditions such that the nuclease(s) are expressed;
30 culturing the cells over a period of time; and measuring the growth of cells in culture at various time intervals. In certain embodiments, the growth of the cells is determined by spectrophotometry, for example by determining the optical density (OD) of the cultured cells at a suitable wavelength (*e.g.*, OD_{600 nm}). The time intervals at which cell growth is determined may be, for example, hours or days (*e.g.*,

2 days, 3 days, 4 days, 5 days, 6, days, 7 days, 8 days, 9 days, 10 days, or even longer) after introduction (or induction) of the nuclease expression cassettes. The nuclease may comprises a ZFN, a ZFN pair, a meganuclease with an engineered DNA-binding domain or a fusion of a naturally-occurring or engineered meganuclease DNA-binding domain and a heterologous cleavage domain. Furthermore, the methods can be performed in a host cell comprising the target sequence recognized by the nuclease (*e.g.*, a reporter construct as described herein). Alternatively, the methods may be performed in a host cell that does not contain the target sequence recognized by the nuclease, as a toxic nuclease will delay yeast growth in the presence or absence of its target sequence.

[0015] In another aspect, methods of selecting a biologically active nuclease (*e.g.*, ZFN, ZFN pair or homing nuclease) are provided. The methods comprise determining nucleases or that cleaves at a selected target site by any of the methods described herein; and determining the toxicity of the nuclease(s) using any of the methods described herein, wherein biologically active nuclease(s) exhibiting cleavage activity and low toxicity are selected.

[0016] In any of the methods described herein, levels of reporter gene activity may be measured directly, for example by directly assaying the levels of the reporter gene product (*e.g.*, GFP fluorescence). Alternatively, levels of the reporter gene can be assayed by measuring the levels of a downstream product (*e.g.*, enzymatic product) of the reaction that requires function of the protein encoded by the reporter gene. In addition, in any of these methods, expression of the nuclease(s) may be driven by a constitutive or inducible promoter. Furthermore, in any of the methods described herein, the nuclease(s) (*e.g.*, ZFN, ZFN pair, engineered homing endonuclease and/or fusion or a naturally occurring or engineered homing endonuclease DNA-binding domain and heterologous cleavage domain) may be known to recognize the target sequence, for example from results obtained from in vitro assay experiments.

BRIEF DESCRIPTION OF THE DRAWINGS

[0017] **Figure 1** is a schematic depicting detection of ZFN activity using a single stranded annealing (SSA)-based reporter system. “P_{GALI}” refers to a *GALI* promoter driving expression of a zinc finger nuclease (ZFN1 or ZFN2); “CYC1t” refers to a CYC1 transcription terminator; “*HIS3*” refers to a wild-type yeast gene *HIS3* which complements specific auxotrophic mutations in yeast (His- phenotype);

“*LEU2*” refers to a wild-type yeast gene *LEU2* which complements specific auxotrophic mutations in yeast (Leu- phenotype); “*HO-L*” refers to the left homology arm of the reporter construct which targets the reporter to the *HO* locus; “*HO-R*” refers to the right homology arm of the reporter construct which targets the reporter to the *HO* locus; “*P_{PGK1}*” refers to a portion of a *PGK1* promoter; “*MEL*” and “*EL1*” refer to a sequence, which when operably linked, encodes a functional Mel1 enzyme; “target” refers to a sequence containing target site(s) for the ZFNs ; “*KanMX*” refers to a sequence encoding kanamycin resistance; “*ChrIV*” refers to chromosome IV; “*DSB*” refers to double stranded break processing; “*SSA*” refers to single strand annealing.

[0018] Figure 2 is a schematic depicting an exemplary *SSA MEL1* reporter construct.

[0019] Figure 3 shows *in vitro* binding data obtained for various ZFNs targeted to the *NME1* locus.

[0020] Figure 4, panels A and B, are graphs showing results from *SSA* annealing assays and toxicity studies for *NME-ZFNs*. Fig. 4A shows *MEL1* activity of the ZFN pairs shown on the x-axis in yeast cells containing the *MEL1* reporter construct with inserted *NME1* target sequence. The left-most bar shows *MEL1* activity in the yeast cells prior to induction of the ZFN expression with galactose with the indicated ZFN pairs; the bar 2nd from the left shows *MEL1* activity in the yeast cells 2 hours after induction of expression of the indicated ZFN pairs; the bar 2nd from the right shows *MEL1* activity in the yeast cells 4 hours after induction of expression of the indicated ZFN pairs; and the right-most bar shows *MEL1* activity in the yeast cells 6 hours after induction of expression of the indicated ZFN pairs.

[0021] Fig. 4B depicts growth, as measured by spectrophotometry at OD₆₀₀, of yeast host cells containing the *MEL1* reporter constructs containing the *NME1* target sequence at various times after introduction of the *NME1*-targeted ZFN pairs indicated on the x-axis. The left-most bar shows OD₆₀₀ of the yeast cells prior to transfection with the indicated ZFN pairs; the bar 2nd from the left shows OD₆₀₀ of the yeast cells 23 hours after introduction of the indicated ZFN pairs; the bar 2nd from the right shows OD₆₀₀ of the yeast cells 27 hours after introduction of the indicated ZFN pairs; and the right-most bar shows OD₆₀₀ of the yeast cells 30 hours after introduction of the indicated ZFN pairs.

[0022] **Figure 5** is a blot showing activity of selected NME1-targeted ZFN pairs in human K562 cells. The percent of non-homologous end joining (NHEJ) is shown below each lane.

[0023] **Figure 6** is a blot depicting activity of NME-1 targeted ZFN pair 13674 and 13677 in human K562 cells. “GFP” refers to the green fluorescent protein negative control; “D2” refers to activity 2 days after introduction of the ZFN pair; “D9” refers to activity 9 days after introduction of the ZFN pair; and “+” refers to the positive control. The percent signal is indicated below each lane.

[0024] **Figure 7**, panels A and B, are graphs showing results from SSA annealing assays and toxicity studies for PD1-ZFNs. Fig. 7A shows MEL1 activity of the ZFN pairs shown on the x-axis in yeast cells containing the MEL1 reporter construct with inserted PD1 target sequence. The left bar shows MEL1 activity in the yeast cells prior to induction of expression of the indicated ZFN pairs and the right bar shows MEL1 activity in the yeast cells 6 hours after induction of expression of the indicated ZFN pairs.

[0025] Fig. 7B depicts growth, as measured by spectrophotometry at OD₆₀₀, of yeast host cells containing the MEL1 reporter constructs containing the PD1 target sequence at various times after introduction of the NME1-targeted ZFN pairs indicated on the x-axis. The left bar shows OD₆₀₀ of the yeast cells prior to transfection with the indicated ZFN pairs and the right bar shows OD₆₀₀ of the yeast cells 30 hours after introduction of the indicated ZFN pairs.

[0026] **Fig. 8** is a schematic depicting an exemplary SSA *MEL1* counter-selectable SSA reporter construct.

[0027] **Fig. 9**, panels A and B, are graphs showing results from SSA annealing assays and toxicity studies for ZFNs targeted to the *golden* gene of zebrafish. Fig. 9A shows MEL1 activity of the ZFN pairs shown on the x-axis in yeast cells containing the MEL1 reporter construct with inserted *golden* target sequence. The left-most bar shows MEL1 activity in the yeast cells prior to induction of the ZFN expression with galactose with the indicated ZFN pairs; the bar 2nd from the left shows MEL1 activity in the yeast cells 2 hours after induction of expression of the indicated ZFN pairs; the bar 2nd from the right shows MEL1 activity in the yeast cells 4 hours after induction of expression of the indicated ZFN pairs; and the right-most bar shows MEL1 activity in the yeast cells 6 hours after induction of expression of the indicated ZFN pairs.

[0028] Fig. 9B depicts growth, as measured by spectrophotometry at OD₆₀₀, of yeast host cells containing the MEL1 reporter constructs containing the zebrafish *golden* target sequence at various times after introduction of the *golden*-targeted ZFN pairs indicated on the x-axis. The left-most bar shows OD₆₀₀ of the yeast cells prior to transfection with the indicated ZFN pairs; the bar 2nd from the left shows OD₆₀₀ of the yeast cells 23 hours after introduction of the indicated ZFN pairs; the bar 2nd from the right shows OD₆₀₀ of the yeast cells 27 hours after introduction of the indicated ZFN pairs; and the right-most bar shows OD₆₀₀ of the yeast cells 30 hours after introduction of the indicated ZFN pairs.

[0029] Fig. 10, panels A and B, are graphs showing results from SSA annealing assays and toxicity studies for ZFNs targeted to the *notail* gene of zebrafish. Fig. 10A shows MEL1 activity of the ZFN pairs shown on the x-axis in yeast cells containing the MEL1 reporter construct with inserted *notail* target sequence. The left-most bar shows MEL1 activity in the yeast cells prior to induction of the ZFN expression with galactose with the indicated ZFN pairs; the bar 2nd from the left shows MEL1 activity in the yeast cells 2 hours after induction of expression of the indicated ZFN pairs; the bar 2nd from the right shows MEL1 activity in the yeast cells 4 hours after induction of expression of the indicated ZFN pairs; and the right-most bar shows MEL1 activity in the yeast cells 6 hours after induction of expression of the indicated ZFN pairs.

[0030] Fig. 10B depicts growth, as measured by spectrophotometry at OD₆₀₀, of yeast host cells containing the MEL1 reporter constructs containing the zebrafish *notail* target sequence at various times after introduction of the *notail*-targeted ZFN pairs indicated on the x-axis. The left-most bar shows OD₆₀₀ of the yeast cells prior to transfection with the indicated ZFN pairs; the bar 2nd from the left shows OD₆₀₀ of the yeast cells 23 hours after introduction of the indicated ZFN pairs; the bar 2nd from the right shows OD₆₀₀ of the yeast cells 27 hours after introduction of the indicated ZFN pairs; and the right-most bar shows OD₆₀₀ of the yeast cells 30 hours after introduction of the indicated ZFN pairs.

[0031] Figure 11 shows pigmentation of zebrafish embryos upon disruption of the *golden* gene. The top panel shows a wild-type organism. The second panel from the top shows a zebrafish embryo when the *golden* gene was mutated as described in Lamason et al. (2005) *Science* 310(5755):1782-6. The left most bottom panel shows eye pigmentation in zebrafish with a *gol*^{bl+/-} background. The 3 right

bottom panels show eye pigmentation in $gol^{bl+/-}$ zebrafish injected with 5 ng of ZFN mRNA directed against golden gene.

[0032] Figure 12, panels A to D, show tail formation of zebrafish embryos upon disruption of the *notail/Brachyury (ntl)* gene. Fig. 12A shows a wild-type zebrafish embryo. Fig 12B shows a zebrafish embryo when the *notail* gene was mutated as described in Amacher et al. (2002) *Development* 129(14):3311-23. Fig. 12C shows a zebrafish embryo with $ntl^{+/-}$ genotype and Fig. 12D shows a zebrafish embryo with a $ntl^{+/-}$ genotype injected with 5 ng of ZFN mRNA directed against *notail* gene.

[0033] Figure 13 is a graph showing results of growth assays of yeast reporter strains expressing various ZFN constructs following selection in counterselection medium (5-FOA) and negative selection in ura- media. Yeast cells were transformed with either an empty expression vector ("vector"), a ZFN ("8266") or a pool of the same ZFN with five different linker sequences ("pool"). The left part of the graph shows the growth in the presence of 5-FOA. The right part of the graph shows growth of the yeast cells in absence of uracil. The bars over each label shows growth after the indicated periods of ZFN induction. The left bar ($t=0$) shows growth with no ZFN induction; the middle bar shows growth when ZFNs were induced for 6 hours ($t=6$) and the right bar shows when ZFNs were induced for 24 hours ($t=24$).

DETAILED DESCRIPTION

[0034] Described herein are compositions and methods for high throughput *in vivo* screening systems for identifying functional nucleases. In particular, the assays use a reporter system to monitor the ability of a nuclease to induce a double-stranded break at their target site. In addition, the assays can be used to determine the effect of the nuclease on cell growth (toxicity).

[0035] Engineered nuclease technology is based on the engineering of naturally occurring DNA-binding proteins. For example, engineering of homing endonucleases with tailored DNA-binding specificities has been described. Chames et al. (2005) *Nucleic Acids Res* 33(20):e178; Arnould et al. (2006) *J. Mol. Biol.* 355:443-458. In addition, engineering of ZFPs has also been described. See, e.g., U.S. Patent Nos. 6,534,261; 6,607,882; 6,824,978; 6,979,539; 6,933,113; 7,163,824; and 7,013,219.

[0036] In addition, ZFPs have been attached to nuclease domains to create ZFNs – a functional entity that is able to recognize its intended gene target through its engineered (ZFP) DNA binding domain and the nuclease causes the gene to be cut near the ZFP binding site. See, e.g., Kim et al. (1996) *Proc Natl Acad Sci USA* 93(3):1156-1160. More recently, ZFNs have been used for genome modification in a variety of organisms. See, for example, United States Patent Publications 20030232410; 20050208489; 20050026157; 20050064474; 20060188987; 20060063231; and International Publication WO 07/014275.

[0037] Although the rules that allow engineering of ZFPs to bind to specific DNA sequences are well characterized and accurately identify specific ZFPs, these same ZFPs may not bind with equal affinity and/or specificity when incorporated into a ZFN. For example, it is likely that the chromosomal substrate can affect the precise dimerization of nuclease domains in living cells, consequently diminishing the cleavage potential, and that the precise chromatin architecture over a given genomic locus will differentially affect the ability of ZFNs to bind and cleave their intended target sequence. In addition, it is difficult if not impossible for *in vitro* assays to mimic the search parameters that a designed DNA binding domain is subjected to when presented with a cellular genome in chromatinized form. As a result, it is essential to test numerous variants in the relevant organism, or cell lineage, to identify a ZFN displaying the optimal characteristics for gene modification.

[0038] Furthermore, since every *in vivo* system has its own peculiarities, it is necessary to develop specific detection assays to determine ZFN action. Thus, unlike previously described *in vivo* screening methods which screen for homing endonucleases with binding specificity different from the naturally occurring homing endonuclease, the methods described herein provide a rapid and efficient way of ranking nucleases already known to bind to a particular target site by predicting their *in vivo* functionality as well as the toxicity of a nuclease to the host cell.

[0039] Thus, the methods and compositions described herein provide highly efficient and rapid methods for identifying nucleases that are biologically active *in vivo*. In addition to accurately predicting *in vivo* nucleases functionality, the assays described herein also can be used to determine nuclease toxicity, thereby allowing identification of the safest and most functionally active proteins.

General

[0040] Practice of the methods, as well as preparation and use of the compositions disclosed herein employ, unless otherwise indicated, conventional techniques in molecular biology, biochemistry, chromatin structure and analysis, computational chemistry, cell culture, recombinant DNA and related fields as are within the skill of the art. These techniques are fully explained in the literature. *See*, for example, Sambrook *et al.* MOLECULAR CLONING: A LABORATORY MANUAL, Second edition, Cold Spring Harbor Laboratory Press, 1989 and Third edition, 2001; Ausubel *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, 1987 and periodic updates; the series METHODS IN ENZYMOLOGY, Academic Press, San Diego; Wolffe, CHROMATIN STRUCTURE AND FUNCTION, Third edition, Academic Press, San Diego, 1998; METHODS IN ENZYMOLOGY, Vol. 304, "Chromatin" (P.M. Wassarman and A. P. Wolffe, eds.), Academic Press, San Diego, 1999; and METHODS IN MOLECULAR BIOLOGY, Vol. 119, "Chromatin Protocols" (P.B. Becker, ed.) Humana Press, Totowa, 1999.

Definitions

[0041] The terms "nucleic acid," "polynucleotide," and "oligonucleotide" are used interchangeably and refer to a deoxyribonucleotide or ribonucleotide polymer, in linear or circular conformation, and in either single- or double-stranded form. For the purposes of the present disclosure, these terms are not to be construed as limiting with respect to the length of a polymer. The terms can encompass known analogues of natural nucleotides, as well as nucleotides that are modified in the base, sugar and/or phosphate moieties (*e.g.*, phosphorothioate backbones). In general, an analogue of a particular nucleotide has the same base-pairing specificity; *i.e.*, an analogue of A will base-pair with T.

[0042] The terms "polypeptide," "peptide" and "protein" are used interchangeably to refer to a polymer of amino acid residues. The term also applies to amino acid polymers in which one or more amino acids are chemical analogues or modified derivatives of a corresponding naturally-occurring amino acids.

[0043] "Binding" refers to a sequence-specific, non-covalent interaction between macromolecules (*e.g.*, between a protein and a nucleic acid). Not all components of a binding interaction need be sequence-specific (*e.g.*, contacts with phosphate residues in a DNA backbone), as long as the interaction as a whole is

sequence-specific. Such interactions are generally characterized by a dissociation constant (K_d) of 10^{-6} M⁻¹ or lower. "Affinity" refers to the strength of binding: increased binding affinity being correlated with a lower K_d .

[0044] A "binding protein" is a protein that is able to bind non-covalently to another molecule. A binding protein can bind to, for example, a DNA molecule (a DNA-binding protein), an RNA molecule (an RNA-binding protein) and/or a protein molecule (a protein-binding protein). In the case of a protein-binding protein, it can bind to itself (to form homodimers, homotrimers, *etc.*) and/or it can bind to one or more molecules of a different protein or proteins. A binding protein can have more than one type of binding activity. For example, zinc finger proteins have DNA-binding, RNA-binding and protein-binding activity.

[0045] A "zinc finger DNA binding protein" (or binding domain) is a protein, or a domain within a larger protein, that binds DNA in a sequence-specific manner through one or more zinc fingers, which are regions of amino acid sequence within the binding domain whose structure is stabilized through coordination of a zinc ion. The term zinc finger DNA binding protein is often abbreviated as zinc finger protein or ZFP.

[0046] Zinc finger binding domains can be "engineered" to bind to a predetermined nucleotide sequence. Non-limiting examples of methods for engineering zinc finger proteins are design and selection. A designed zinc finger protein is a protein not occurring in nature whose design/composition results principally from rational criteria. Rational criteria for design include application of substitution rules and computerized algorithms for processing information in a database storing information of existing ZFP designs and binding data. See, for example, US Patents 6,140,081; 6,453,242; and 6,534,261; see also WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 and WO 03/016496.

[0047] A "selected" zinc finger protein is a protein not found in nature whose production results primarily from an empirical process such as phage display, interaction trap or hybrid selection. See *e.g.*, US 5,789,538; US 5,925,523; US 6,007,988; US 6,013,453; US 6,200,759; WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970 WO 01/88197 and WO 02/099084.

[0048] "Cleavage" refers to the breakage of the covalent backbone of a DNA molecule. Cleavage can be initiated by a variety of methods including, but not limited to, enzymatic or chemical hydrolysis of a phosphodiester bond. Both single-stranded cleavage and double-stranded cleavage are possible, and double-stranded cleavage

can occur as a result of two distinct single-stranded cleavage events. DNA cleavage can result in the production of either blunt ends or staggered ends. In certain embodiments, fusion polypeptides are used for targeted double-stranded DNA cleavage.

5 [0049] An "cleavage half-domain" is a polypeptide sequence which, in conjunction with a second polypeptide (either identical or different) forms a complex having cleavage activity (preferably double-strand cleavage activity). The terms "first and second cleavage half-domains;" "+ and – cleavage half-domains" and "right and left cleavage half-domains" are used interchangeably to refer to pairs of cleavage half-
10 domains that dimerize.

[0050] An "engineered cleavage half-domain" is a cleavage half-domain that has been modified so as to form obligate heterodimers with another cleavage half-domain (*e.g.*, another engineered cleavage half-domain). *See, also*, U.S. Patent Application Nos. 10/912,932 and 11/304,981 and U.S. Provisional Application No.
15 60/808,486 (filed May 25, 2006), incorporated herein by reference in their entireties.

[0051] "Chromatin" is the nucleoprotein structure comprising the cellular genome. Cellular chromatin comprises nucleic acid, primarily DNA, and protein, including histones and non-histone chromosomal proteins. The majority of eukaryotic cellular chromatin exists in the form of nucleosomes, wherein a
20 nucleosome core comprises approximately 150 base pairs of DNA associated with an octamer comprising two each of histones H2A, H2B, H3 and H4; and linker DNA (of variable length depending on the organism) extends between nucleosome cores. A molecule of histone H1 is generally associated with the linker DNA. For the purposes of the present disclosure, the term "chromatin" is meant to encompass all types of
25 cellular nucleoprotein, both prokaryotic and eukaryotic. Cellular chromatin includes both chromosomal and episomal chromatin.

[0052] A "chromosome," is a chromatin complex comprising all or a portion of the genome of a cell. The genome of a cell is often characterized by its karyotype, which is the collection of all the chromosomes that comprise the genome of the cell.
30 The genome of a cell can comprise one or more chromosomes.

[0053] An "episome" is a replicating nucleic acid, nucleoprotein complex or other structure comprising a nucleic acid that is not part of the chromosomal karyotype of a cell. Examples of episomes include plasmids and certain viral genomes.

[0054] A "target site" or "target sequence" is a nucleic acid sequence that defines a portion of a nucleic acid to which a binding molecule will bind, provided sufficient conditions for binding exist. For example, the sequence 5'-GAATTC-3' is a target site for the Eco RI restriction endonuclease.

5 [0055] An "exogenous" molecule is a molecule that is not normally present in a cell, but can be introduced into a cell by one or more genetic, biochemical or other methods. "Normal presence in the cell" is determined with respect to the particular developmental stage and environmental conditions of the cell. Thus, for example, a molecule that is present only during embryonic development of muscle is an
10 exogenous molecule with respect to an adult muscle cell. Similarly, a molecule induced by heat shock is an exogenous molecule with respect to a non-heat-shocked cell. An exogenous molecule can comprise, for example, a functioning version of a malfunctioning endogenous molecule or a malfunctioning version of a normally-functioning endogenous molecule.

15 [0056] An exogenous molecule can be, among other things, a small molecule, such as is generated by a combinatorial chemistry process, or a macromolecule such as a protein, nucleic acid, carbohydrate, lipid, glycoprotein, lipoprotein, polysaccharide, any modified derivative of the above molecules, or any complex comprising one or more of the above molecules. Nucleic acids include DNA and
20 RNA, can be single- or double-stranded; can be linear, branched or circular; and can be of any length. Nucleic acids include those capable of forming duplexes, as well as triplex-forming nucleic acids. See, for example, U.S. Patent Nos. 5,176,996 and 5,422,251. Proteins include, but are not limited to, DNA-binding proteins, transcription factors, chromatin remodeling factors, methylated DNA binding
25 proteins, polymerases, methylases, demethylases, acetylases, deacetylases, kinases, phosphatases, integrases, recombinases, ligases, topoisomerases, gyrases and helicases.

[0057] An exogenous molecule can be the same type of molecule as an endogenous molecule, *e.g.*, an exogenous protein or nucleic acid. For example, an
30 exogenous nucleic acid can comprise an infecting viral genome, a plasmid or episome introduced into a cell, or a chromosome that is not normally present in the cell. Methods for the introduction of exogenous molecules into cells are known to those of skill in the art and include, but are not limited to, lipid-mediated transfer (*i.e.*, liposomes, including neutral and cationic lipids), electroporation, direct injection, cell

fusion, particle bombardment, calcium phosphate co-precipitation, DEAE-dextran-mediated transfer and viral vector-mediated transfer.

[0058] By contrast, an "endogenous" molecule is one that is normally present in a particular cell at a particular developmental stage under particular environmental conditions. For example, an endogenous nucleic acid can comprise a chromosome, the genome of a mitochondrion, chloroplast or other organelle, or a naturally-occurring episomal nucleic acid. Additional endogenous molecules can include proteins, for example, transcription factors and enzymes.

[0059] A "fusion" molecule is a molecule in which two or more subunit molecules are linked, preferably covalently. The subunit molecules can be the same chemical type of molecule, or can be different chemical types of molecules.

Examples of the first type of fusion molecule include, but are not limited to, fusion proteins (for example, a fusion between a ZFP DNA-binding domain and a cleavage domain) and fusion nucleic acids (for example, a nucleic acid encoding the fusion protein described *supra*). Examples of the second type of fusion molecule include, but are not limited to, a fusion between a triplex-forming nucleic acid and a polypeptide, and a fusion between a minor groove binder and a nucleic acid.

[0060] Expression of a fusion protein in a cell can result from delivery of the fusion protein to the cell or by delivery of a polynucleotide encoding the fusion protein to a cell, wherein the polynucleotide is transcribed, and the transcript is translated, to generate the fusion protein. Trans-splicing, polypeptide cleavage and polypeptide ligation can also be involved in expression of a protein in a cell. Methods for polynucleotide and polypeptide delivery to cells are presented elsewhere in this disclosure.

[0061] "Eukaryotic" cells include, but are not limited to, fungal cells (such as yeast), plant cells, animal cells, mammalian cells and human cells (*e.g.*, T-cells).

[0062] The terms "operative linkage" and "operatively linked" (or "operably linked") are used interchangeably with reference to a juxtaposition of two or more components (such as sequence elements), in which the components are arranged such that both components function normally and allow the possibility that at least one of the components can mediate a function that is exerted upon at least one of the other components. By way of illustration, a transcriptional regulatory sequence, such as a promoter, is operatively linked to a coding sequence if the transcriptional regulatory sequence controls the level of transcription of the coding sequence in response to the

presence or absence of one or more transcriptional regulatory factors. A transcriptional regulatory sequence is generally operatively linked in *cis* with a coding sequence, but need not be directly adjacent to it. For example, an enhancer is a transcriptional regulatory sequence that is operatively linked to a coding sequence, even though they are not contiguous.

[0063] With respect to fusion polypeptides, the term "operatively linked" can refer to the fact that each of the components performs the same function in linkage to the other component as it would if it were not so linked. For example, with respect to a fusion polypeptide in which a ZFP DNA-binding domain is fused to a cleavage domain, the ZFP DNA-binding domain and the cleavage domain are in operative linkage if, in the fusion polypeptide, the ZFP DNA-binding domain portion is able to bind its target site and/or its binding site, while the cleavage domain is able to cleave DNA in the vicinity of the target site.

[0064] A "vector" is capable of transferring gene sequences to target cells. Typically, "vector construct," "expression vector," and "gene transfer vector," mean any nucleic acid construct capable of directing the expression of a gene of interest and which can transfer gene sequences to target cells. Thus, the term includes cloning, and expression vehicles, as well as integrating vectors.

[0065] A "reporter gene" or "reporter sequence" refers to any sequence that produces a protein product that is easily measured, preferably in a routine assay. Suitable reporter genes include, but are not limited to, Mel1, chloramphenicol acetyl transferase (CAT), light generating proteins, and β -galactosidase.

Overview

[0066] Described herein are compositions and methods for the identification of nucleases that cleave their target sites with the highest frequency and are not toxic to the host cell. Reporter constructs comprising a target site for the nucleases to be tested are described as are host cells comprising these reporter constructs. In the methods described herein, the reporter construct comprising the target site for the nuclease(s) is introduced into a host cell (e.g., yeast cell) to create a reporter strain. When the nuclease(s) are expressed in the cell and induce a double stranded break (DSB) at their target site (e.g., induce a double-stranded break), the reporter gene is reconstituted by the host cell's single-stranded annealing (SSA) machinery. Expression of the reporter gene is readily determined by standard techniques and the

levels of reporter gene expression reflect the ability of the nuclease to cleave at the target site. In addition, the host cells can be readily assayed to determine the effect of nuclease expression on cell growth.

[0067] Thus, described herein are rapid and efficient high throughput screening methods for determining the most active and least toxic nucleases from a panel of nucleases known to bind to a particular target site. In addition to allowing ranking of nucleases according to their activity at the target locus, the present disclosure also allows for a determination as to which nucleases display non-specific cutting of the genome.

[0068] The reagents and methods described herein that allow for *in vivo* characterization of nuclease action can be conducted in budding yeast. The rapid and versatile genetics of yeast allows testing of a large panel of nucleases in a simple assay in high throughput fashion. The reagents and systems can be used to screen nucleases designed against any gene from any organism and the disclosure has been validated as correctly identifying the optimally active nuclease pairs using lower vertebrate, plant, and human cell cultures.

Reporter Constructs

[0069] The methods and systems described herein make use of a reporter constructs comprising a sequence containing a target sequence for the nucleases to be tested. The reporter construct is designed so that the reporter gene is functional only when the target sequence is cleaved and the reporter reconstituted by single-strand annealing (SSA) of the reporter gene sequences. Typically, a reporter construct is generated such that any nuclease target sequence(s) can be readily inserted into the middle of the reporter gene sequence, for example via a polylinker (see, Figs. 1 and 2).

[0070] One or more target sites for the nuclease(s) to be screened can be inserted into the reporter constructs by any suitable methodology, including PCR or commercially available cloning systems such as TOPO® and/or Gateway® cloning systems. In certain embodiments, the target site comprises a concatamer of target sites. See, also, Example 1. Target sites can be from prokaryotic or eukaryotic genes, for example, mammalian (*e.g.*, human), yeast or plant cells.

[0071] Any reporter gene that provides a detectable signal can be used, including but not limited, enzymes that catalyze the production of a detectable

product (e.g. proteases, nucleases, lipases, phosphatases, sugar hydrolases and esterases). Non-limiting examples of suitable reporter genes that encode enzymes include, for example, MEL1, CAT (chloramphenicol acetyl transferase; Alton and Vapnek (1979) *Nature* 282:864-869), luciferase, β -galactosidase, β -glucuronidase, β -lactamase, horseradish peroxidase and alkaline phosphatase (e.g., Toh, et al. (1980) *Eur. J. Biochem.* 182:231-238; and Hall et al. (1983) *J. Mol. Appl. Gen.* 2:101).

Reporter genes that provide a detectable signal directly may also be employed, for example, fluorescent proteins such as, for example, GFP (green fluorescent protein). Fluorescence is detected using a variety of commercially available fluorescent detection systems, including a fluorescence-activated cell sorter (FACS) system for example.

[0072] In certain embodiments, the reporter gene encodes an enzyme, for example MEL1. The use of the secreted MEL1 reporter gene allows for convenient detection of recombination events directly from the growth media without requirement for cell lysis, as compared to the classic β -galactosidase yeast reporter gene (Aho et al. (1997) *Anal Biochem* 253:270-272).

[0073] As shown in Figs. 1 and 2, the reporter construct also typically comprises sequences flanking the reporter-target-reporter sequences that are homologous to regions of the host cell genomic DNA. These "homology arms" allow for targeted integration of the reporter construct into the host cell to generate a stable reporter host cell line. The homology arms can be to any genomic sequence of the host cell. Preferably, the homology arms direct insertion of the reporter construct to a non-essential site in the host cell genome, for example the HO locus in yeast. Other non-limiting examples of suitable insertion sites include auxotrophy markers such as *URA3*, *LYS2*, and *TRP1*. Preferably, the reporter construct is inserted into a locus whose mutation (or knockout) does not quantitatively affect host cell growth. The reporter constructs may be integrated into the host cell genome using standard techniques. See, e.g., Chames et al., supra and Arnould et al., supra. Alternatively, the reporter constructs can be maintained episomally.

[0074] The reporter constructs may also comprise one or more selectable markers. Positive selection markers are those polynucleotides that encode a product that enables only cells that carry and express the gene to survive and/or grow under certain conditions. For example, cells that express antibiotic resistance genes (e.g. Kan^r or Neo^r) gene are resistant to the antibiotics or their analogs (e.g. G418), while

cells that do not express these resistance genes are killed in the presence of antibiotics. Other examples of positive selection markers including hygromycin resistance, Zeocin™ resistance and the like will be known to those of skill in the art (see, Golstein and McCusker (1999) *Yeast* 15:1541-1553). Negative selection markers are those polynucleotides that encode a produce that enables only cells that carry and express the gene to be killed under certain conditions. For example, cells that express thymidine kinase (e.g., herpes simplex virus thymidine kinase, HSV-TK) are killed when gancyclovir is added. Other negative selection markers are known to those skilled in the art. The selectable marker need not be a transgene and, additionally, reporters and selectable markers can be used in various combinations.

[0075] The reporter construct may also include additional reporter genes, for example genes that reflect off-target nuclease activity by indicating that the cell is undergoing a DNA damage response (DDR). Non-limiting examples of such suitable additional off-target reporters include genes known to be upregulated by induction of even a single DSB, for example *RNR2*, *RNR4*, *DIN7*, *PCL5*, *DUN1*. See, also, Lee et al. (2000) *Cold Spring Harb Symp Quant Biol.* 65:303:314. Additional reporters can be independently introduced and may be transiently expressed or stably integrated into the host cell.

Host Cells

[0076] Any host cell that reconstitutes a functional reporter upon cleavage of the target sequence by the nuclease(s) can be used in the practice of the present disclosure. The cell types can be cell lines or natural (e.g., isolated) cells such as, for example, primary cells. Cell lines are available, for example from the American Type Culture Collection (ATCC), or can be generated by methods known in the art, as described for example in Freshney et al., *Culture of Animal Cells, A Manual of Basic Technique*, 3rd ed., 1994, and references cited therein. Similarly, cells can be isolated by methods known in the art. Other non-limiting examples of cell types include cells that have or are subject to pathologies, such as cancerous cells and transformed cells, pathogenically infected cells, stem cells, fully differentiated cells, partially differentiated cells, immortalized cells and the like. Prokaryotic (e.g., bacterial) or eukaryotic (e.g., yeast, plant, fungal, piscine and mammalian cells such as feline, canine, murine, bovine, porcine and human) cells can be used, with eukaryotic cells

being preferred. Suitable mammalian cell lines include CHO (Chinese hamster ovary) cells, HEP-G2 cells, BaF-3 cells, Schneider cells, COS cells (monkey kidney cells expressing SV40 T-antigen), CV-1 cells, HuTu80 cells, NTERA2 cells, NB4 cells, HL-60 cells and HeLa cells, 293 cells (see, e.g., Graham et al. (1977) *J. Gen. Virol.* 36:59), and myeloma cells like SP2 or NS0 (see, e.g., Galfre and Milstein (1981) *Meth. Enzymol.* 73(B):3 46. Other eukaryotic cells include, for example, insect (e.g., *sp. frugiperda*), fungal cells, including yeast (e.g., *S. cerevisiae*, *S. pombe*, *P. pastoris*, *K. lactis*, *H. polymorpha*), and plant cells (Fleer, R. (1992) *Current Opinion in Biotechnology* 3:486 496).

[0077] In a preferred embodiment, the host cell is a yeast cell. Yeast cells are advantageously employed because the deletion of the intervening sequences required for the reconstitution of the reporter is an efficient process in these cells and permits the scanning of large genomic targets. Yeast cells survive the introduction of a DSB even if the target is up to 25 kb (Vaze et al. (2002) *Mol Cell* 10:373-385). In addition, as long as 400 base pairs of homologous regions within the reporter construct are provided, 100% of yeast cells survive to the break using the SSA repair pathway (Sugawara et al. (2000) *Mol Cell Biol* 20:5300-5309). Any strain of yeast cell can be used, including, by way of example, 69-1B or BY4741.

20 Nucleases

[0078] The methods and compositions described herein are broadly applicable and may involve any nuclease of interest. Non-limiting examples of nucleases include meganucleases and zinc finger nucleases. The nuclease may comprise heterologous DNA-binding and cleavage domains (e.g., zinc finger nucleases; meganuclease DNA-binding domains with heterologous cleavage domains) or, alternatively, the DNA-binding domain of a naturally-occurring nuclease may be altered to bind to a selected target site (e.g., a meganuclease that has been engineered to bind to site different than the cognate binding site).

[0079] In certain embodiment, the nuclease is a meganuclease (homing endonuclease). Naturally-occurring meganucleases recognize 15-40 base-pair cleavage sites and are commonly grouped into four families: the LAGLIDADG family, the GIY-YIG family, the His-Cyst box family and the HNH family. Exemplary homing endonucleases include I-SceI, I-CeuI, PI-PspI, PI-Sce, I-SceIV, I-CsmI, I-PanI, I-SceII, I-PpoI, I-SceIII, I-CreI, I-TevI, I-TevII and I-TevIII. Their

recognition sequences are known. See also U.S. Patent No. 5,420,032; U.S. Patent No. 6,833,252; Belfort *et al.* (1997) *Nucleic Acids Res.* **25**:3379–3388; Dujon *et al.* (1989) *Gene* **82**:115–118; Perler *et al.* (1994) *Nucleic Acids Res.* **22**, 1125–1127; Jasin (1996) *Trends Genet.* **12**:224–228; Gimble *et al.* (1996) *J. Mol. Biol.* **263**:163–180; Argast *et al.* (1998) *J. Mol. Biol.* **280**:345–353 and the New England Biolabs catalogue.

[0080] DNA-binding domains from naturally-occurring meganucleases, primarily from the LAGLIDADG family, have been used to promote site-specific genome modification in plants, yeast, *Drosophila*, mammalian cells and mice, but this approach has been limited to the modification of either homologous genes that conserve the meganuclease recognition sequence (Monet *et al.* (1999), *Biochem. Biophysics. Res. Common.* **255**: 88-93) or to pre-engineered genomes into which a recognition sequence has been introduced (Route *et al.* (1994), *Mol. Cell. Biol.* **14**: 8096-106; Chilton *et al.* (2003), *Plant Physiology.* **133**: 956-65; Puchta *et al.* (1996), *Proc. Natl. Acad. Sci. USA* **93**: 5055-60; Rong *et al.* (2002), *Genes Dev.* **16**: 1568-81; Gouble *et al.* (2006), *J. Gene Med.* **8**(5):616-622). Accordingly, attempts have been made to engineer meganucleases to exhibit novel binding specificity at medically or biotechnologically relevant sites (Porteus *et al.* (2005), *Nat. Biotechnol.* **23**: 967-73; Sussman *et al.* (2004), *J. Mol. Biol.* **342**: 31-41; Epinat *et al.* (2003), *Nucleic Acids Res.* **31**: 2952-62; Chevalier *et al.* (2002) *Molec. Cell* **10**:895-905; Epinat *et al.* (2003) *Nucleic Acids Res.* **31**:2952-2962; Ashworth *et al.* (2006) *Nature* **441**:656-659; Paques *et al.* (2007) *Current Gene Therapy* **7**:49-66; U.S. Patent Publication Nos. 20070117128; 20060206949; 20060153826; 20060078552; and 20040002092). In addition, naturally-occurring or engineered DNA-binding domains from meganucleases have also been operably linked with a cleavage domain from a heterologous nuclease (*e.g.*, FokI).

[0081] In other embodiments, the nuclease is a zinc finger nuclease (ZFN). ZFNs comprise a zinc finger protein that has been engineered to bind to a target site in a gene of choice and cleavage domain or a cleavage half-domain.

[0082] Zinc finger binding domains can be engineered to bind to a sequence of choice. See, for example, Beerli *et al.* (2002) *Nature Biotechnol.* **20**:135-141; Pabo *et al.* (2001) *Ann. Rev. Biochem.* **70**:313-340; Isalan *et al.* (2001) *Nature Biotechnol.* **19**:656-660; Segal *et al.* (2001) *Curr. Opin. Biotechnol.* **12**:632-637; Choo *et al.* (2000) *Curr. Opin. Struct. Biol.* **10**:411-416. An engineered zinc finger binding

domain can have a novel binding specificity, compared to a naturally-occurring zinc finger protein. Engineering methods include, but are not limited to, rational design and various types of selection. Rational design includes, for example, using databases comprising triplet (or quadruplet) nucleotide sequences and individual zinc finger amino acid sequences, in which each triplet or quadruplet nucleotide sequence is associated with one or more amino acid sequences of zinc fingers which bind the particular triplet or quadruplet sequence. See, for example, co-owned U.S. Patents 6,453,242 and 6,534,261, incorporated by reference herein in their entireties.

[0083] Exemplary selection methods, including phage display and two-hybrid systems, are disclosed in US Patents 5,789,538; 5,925,523; 6,007,988; 6,013,453; 6,410,248; 6,140,466; 6,200,759; and 6,242,568; as well as WO 98/37186; WO 98/53057; WO 00/27878; WO 01/88197 and GB 2,338,237. In addition, enhancement of binding specificity for zinc finger binding domains has been described, for example, in co-owned WO 02/077227.

[0084] Selection of target sites; ZFNs and methods for design and construction of fusion proteins (and polynucleotides encoding same) are known to those of skill in the art and described in detail in U.S. Patent Application Publication Nos. 20050064474 and 20060188987, incorporated by reference in their entireties herein.

[0085] In addition, as disclosed in these and other references, zinc finger domains and/or multi-fingered zinc finger proteins may be linked together using any suitable linker sequences, including for example, linkers of 5 or more amino acids in length. See, e.g., U.S. Patent Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The proteins described herein may include any combination of suitable linkers between the individual zinc fingers of the protein.

[0086] Nucleases such as ZFNs and/or meganucleases also comprise a nuclease (cleavage domain, cleavage half-domain). As noted above, the cleavage domain may be heterologous to the DNA-binding domain, for example a zinc finger DNA-binding domain and a cleavage domain from a nuclease or a meganuclease DNA-binding domain and cleavage domain from a different nuclease. Heterologous cleavage domains can be obtained from any endonuclease or exonuclease. Exemplary endonucleases from which a cleavage domain can be derived include, but are not limited to, restriction endonucleases and homing endonucleases. See, for example,

2002-2003 Catalogue, New England Biolabs, Beverly, MA; and Belfort *et al.* (1997) *Nucleic Acids Res.* **25**:3379-3388. Additional enzymes which cleave DNA are known (*e.g.*, S1 Nuclease; mung bean nuclease; pancreatic DNase I; micrococcal nuclease; yeast HO endonuclease; *see also* Linn *et al.* (eds.) *Nucleases*, Cold Spring Harbor Laboratory Press, 1993). One or more of these enzymes (or functional fragments thereof) can be used as a source of cleavage domains and cleavage half-domains.

[0087] Similarly, a cleavage half-domain can be derived from any nuclease or portion thereof, as set forth above, that requires dimerization for cleavage activity. In general, two fusion proteins are required for cleavage if the fusion proteins comprise cleavage half-domains. Alternatively, a single protein comprising two cleavage half-domains can be used. The two cleavage half-domains can be derived from the same endonuclease (or functional fragments thereof), or each cleavage half-domain can be derived from a different endonuclease (or functional fragments thereof). In addition, the target sites for the two fusion proteins are preferably disposed, with respect to each other, such that binding of the two fusion proteins to their respective target sites places the cleavage half-domains in a spatial orientation to each other that allows the cleavage half-domains to form a functional cleavage domain, *e.g.*, by dimerizing. Thus, in certain embodiments, the near edges of the target sites are separated by 5-8 nucleotides or by 15-18 nucleotides. However any integral number of nucleotides or nucleotide pairs can intervene between two target sites (*e.g.*, from 2 to 50 nucleotide pairs or more). In general, the site of cleavage lies between the target sites.

[0088] Restriction endonucleases (restriction enzymes) are present in many species and are capable of sequence-specific binding to DNA (at a recognition site), and cleaving DNA at or near the site of binding. Certain restriction enzymes (*e.g.*, Type IIS) cleave DNA at sites removed from the recognition site and have separable binding and cleavage domains. For example, the Type IIS enzyme *Fok* I catalyzes double-stranded cleavage of DNA, at 9 nucleotides from its recognition site on one strand and 13 nucleotides from its recognition site on the other. See, for example, US Patents 5,356,802; 5,436,150 and 5,487,994; as well as Li *et al.* (1992) *Proc. Natl. Acad. Sci. USA* **89**:4275-4279; Li *et al.* (1993) *Proc. Natl. Acad. Sci. USA* **90**:2764-2768; Kim *et al.* (1994a) *Proc. Natl. Acad. Sci. USA* **91**:883-887; Kim *et al.* (1994b) *J. Biol. Chem.* **269**:31,978-31,982. Thus, in one embodiment, fusion proteins comprise the cleavage domain (or cleavage half-domain) from at least one Type IIS

restriction enzyme and one or more zinc finger binding domains, which may or may not be engineered.

[0089] An exemplary Type IIS restriction enzyme, whose cleavage domain is separable from the binding domain, is *Fok* I. This particular enzyme is active as a dimer. Bitinaite *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95**: 10,570-10,575.

Accordingly, for the purposes of the present disclosure, the portion of the *Fok* I enzyme used in the disclosed fusion proteins is considered a cleavage half-domain. Thus, for targeted double-stranded cleavage and/or targeted replacement of cellular sequences using zinc finger-*Fok* I fusions, two fusion proteins, each comprising a *Fok* I cleavage half-domain, can be used to reconstitute a catalytically active cleavage domain. Alternatively, a single polypeptide molecule containing a zinc finger binding domain and two *Fok* I cleavage half-domains can also be used. Parameters for targeted cleavage and targeted sequence alteration using zinc finger-*Fok* I fusions are provided elsewhere in this disclosure.

[0090] A cleavage domain or cleavage half-domain can be any portion of a protein that retains cleavage activity, or that retains the ability to multimerize (*e.g.*, dimerize) to form a functional cleavage domain.

[0091] Exemplary Type IIS restriction enzymes are described in International Publication WO 07/014275, incorporated herein in its entirety. Additional restriction enzymes also contain separable binding and cleavage domains, and these are contemplated by the present disclosure. See, for example, Roberts *et al.* (2003) *Nucleic Acids Res.* **31**:418-420.

[0092] In certain embodiments, the cleavage domain comprises one or more engineered cleavage half-domain (also referred to as dimerization domain mutants) that minimize or prevent homodimerization, as described, for example, in U.S. Patent Publication Nos. 20050064474 and 20060188987 and in U.S. Application No. 11/805,850 (filed May 23, 2007), the disclosures of all of which are incorporated by reference in their entireties herein. Amino acid residues at positions 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537, and 538 of *Fok* I are all targets for influencing dimerization of the *Fok* I cleavage half-domains.

[0093] Exemplary engineered cleavage half-domains of *Fok* I that form obligate heterodimers include a pair in which a first cleavage half-domain includes mutations at amino acid residues at positions 490 and 538 of *Fok* I and a second cleavage half-domain includes mutations at amino acid residues 486 and 499.

[0094] Thus, in one embodiment, a mutation at 490 replaces Glu (E) with Lys (K); the mutation at 538 replaces Iso (I) with Lys (K); the mutation at 486 replaced Gln (Q) with Glu (E); and the mutation at position 499 replaces Iso (I) with Lys (K). Specifically, the engineered cleavage half-domains described herein were prepared by mutating positions 490 (E→K) and 538 (I→K) in one cleavage half-domain to produce an engineered cleavage half-domain designated “E490K:I538K” and by mutating positions 486 (Q→E) and 499 (I→L) in another cleavage half-domain to produce an engineered cleavage half-domain designated “Q486E:I499L”. The engineered cleavage half-domains described herein are obligate heterodimer mutants in which aberrant cleavage is minimized or abolished. *See, e.g.*, Example 1 of U.S. Provisional Application No. 60/808,486 (filed May 25, 2006), the disclosure of which is incorporated by reference in its entirety for all purposes.

[0095] Engineered cleavage half-domains described herein can be prepared using any suitable method, for example, by site-directed mutagenesis of wild-type cleavage half-domains (*Fok* I) as described in Example 5 of U.S. Patent Publication No. 20050064474 and Example 38 of U.S. Patent Provisional Application Serial No. 60/721,054.

[0096] Nuclease expression constructs can be readily designed using methods known in the art. *See, e.g.*, United States Patent Publications 20030232410; 20050208489; 20050026157; 20050064474; 20060188987; 20060063231; and International Publication WO 07/014275. In certain embodiments, expression of the nuclease is under the control of an inducible promoter, for example the galactokinase promoter which is activated (de-repressed) in the presence of raffinose and/or galactose and repressed in presence of glucose. In particular, the galactokinase promoter is induced and the nuclease(s) expressed upon successive changes in the carbon source (*e.g.*, from glucose to raffinose to galactose). Other non-limiting examples of inducible promoters include *CUP1*, *MET15*, *PHO5*, and tet-responsive promoters.

Identification of Biologically Active Nucleases

[0097] The host cell containing SSA reporter constructs as described herein can be used to identify the most active and the least toxic nucleases from a panel of nucleases engineered to bind to a particular target site. The systems of the present

disclosure take advantage of a particular pathway of homology-directed repair (HDR) called single-strand annealing (SSA). If a double-strand break (DSB) occurs between two flanking homologous regions, repair of the broken chromosome results in a deletion containing a single copy of the repeated sequence (Paques and Haber (1999) *Microbiol Mol Biol Rev* 63:349-404). The engineering of a reporter construct containing two overlapping and non-functional parts of a reporter gene separated by a target sequence permits the easy detection of a nuclease-induced DSB.

[0098] As outlined in Fig. 1, identification of nucleases with the highest in vivo cleavage activity begins with introduction of a reporter construct. The reporter construct can be episomal, for example, using an episome, for example using a yeast centromeric plasmid (YCp). Preferably, the reporter construct is integrated into the genome of the host cell (e.g., yeast), for example by homologous recombination.

[0099] After genotyping the strain for the correct integration of the reporter, the host strain is transformed with nuclease expression vectors. Preferably, nuclease expression is inducible (e.g. galactose-inducible) so that nuclease expression can be induced for a selected amount of time by changing the carbon source in the culture media. After a recovery period required for the cell machinery to repair the induced DSBs, the activity of the reconstituted reporter gene (e.g. Mel1 enzyme) is determined from an aliquot of the media using a suitable (e.g. a colorimetric) assay.

[0100] The activity obtained for each nuclease reflects quantitatively its capacity to induce a DSB within the chromosomal target sequence. The activity is typically normalized to the density of the cells in the culture.

[0101] The *in vivo* screening systems described herein have the added benefit of concomitantly interrogating the entire yeast genome for off-target cleavage by the nuclease(s). The host cells also contain the machinery for a second pathway of DSB repair, namely non-homologous end joining (NHEJ). In the haploid state, yeast cells respond inefficiently to a persistent DSB induced by the continued presence of an endonuclease. Only 0.1% of the cells can survive this type of DSB resulting in a strong delay in growth of the population (Moore and Haber, 1996). Such off-target activity will kill most of the cells and, as cell growth can be easily monitored by spectrophotometric determination of cell density within the culture, nucleases that are least toxic to cells (presumably by virtue of their specificity and lack of off-target cleavage) can be readily identified.

[0102] Furthermore, off-target effects can also be monitored by utilizing a “gain of signal” assay that exploits the DNA damage response (DDR) pathway. Several genes are known to be upregulated by the induction of even a single DSB (Lee et al. (2000) Cold Spring Harb Symp Quant Biol 65:303-314). Accordingly, any
5 of the compositions or methods described herein may further include a second reporter gene under the control of a promoter that is transcribed only in the presence of DSB. Activity of this second reporter gene can be used to detect broadly non-specific nuclease(s).

[0103] A counterselectable gene can also be inserted, for example, in between
10 the interrupted MEL1 gene. This would allow for the selection of active variants from a population of mutated ZFN. Any counterselectable gene can be used, including but not limited to URA3 (Fig. 8). A negative selection can be performed with this gene based on the specific inhibitor, 5-fluoro-orotic acid (FOA) that prevents growth of the prototrophic strains but allows growth of the *ura3* mutants. Ura3-cells
15 (arising from SSA) can be selected on media containing FOA. The URA3+ cells, which contain non active ZFN, are killed because FOA is converted to the toxic compound 5-fluorouracil by the action of decarboxylase, whereas *ura3*- cells are resistant. The negative selection on FOA media is highly discriminating, and usually less than 10⁻² FOA-resistant colonies are Ura+.

[0104] Thus, reporter strains containing this type of counterselectable marker
20 gene are used to eliminate inactive variants, which typically constitute the vast majority of mutants generated in such genetic screens. Cells containing active variants would be resistant to FOA because the URA3 gene will be deleted during the repair of the DSB by SSA. This kind of selection diminishes significantly the work
25 load since most non-functional colonies/variants are eliminated from the screen.

[0105] The following Examples relate to exemplary embodiments of the present disclosure in which the nuclease comprises a ZFN. It will be appreciated that this is for purposes of exemplification only and that other nucleases can be used, for
30 instance homing endonucleases (meganucleases) with engineered DNA-binding domains and/or fusions of naturally occurring or engineered homing endonucleases (meganucleases) DNA-binding domains and heterologous cleavage domains.

EXAMPLES**Example 1: Engineering of a Yeast Reporter Construct**

[0104] A SSA reporter construct (see, Fig. 2) targeted to the HO locus was generated using the yeast integrating plasmid (Yip) HO-poly-KanMX-HO (Voth et al. (2001) *Nucleic Acids Res* 29:E59-59) as follows. A fragment corresponding to nucleotides 1 to 750 of the MEL1 gene (Liljestrom (1985) *Nucleic Acids Res* 13:7257-7268) (relative to the ATG) was cloned into the Sall and BamHI sites of HO-poly-KanMX-HO using the following primers: 5'-aattgtcgacatgtttgctttctactttctaccgc-3' (SEQ ID NO:1) and 5'-aattggatccccccattggagctgcc-3' (SEQ ID NO:2).

Subsequently, a fragment from nucleotides 299 to 2100 were cloned into the SacI and EcoRI sites using the following oligos: 5'-aattgagctcagaccacctgcataataacagc-3' (SEQ ID NO:3) and 5'-aattgaattcgggcaaaaattggtaccaatgc-3' (SEQ ID NO:4). Finally, a 1489 base pair fragment of the PGK1 promoter was cloned into the BsiWI and Sall sites using the following oligos: 5'-Aattcgtagctctaactgatctatccaaaactg-3' (SEQ ID NO:5) and 5'-Aattgtcgacttgatcttttggttttatatttggtg-3' (SEQ ID NO:6).

[0105] MEL1 reporter constructs as described above were further modified to include a Gateway® cassette (Invitrogen), following the manufacturer's instructions. In addition, reporter constructs were generated that included shortened versions of the PGK1 promoter. Constructs lacking the ampicillin resistance gene were also generated.

[0106] Reporter constructs including various ZFN target sites for were also generated. Briefly, one or more copies of the target sites were generated by PCR or by concatamerization and inserted into the Gateway® MEL1 reporter construct using standard molecular biology techniques. Concatamers were constructed electronically using DNAworks (available on the internet) which designed overlapping oligos that tiled across the target site concatamer. The oligos were used to synthesize the synthetic target site concatamer and a two-step cloning process used to introduce the concatamer into the reporter construct. First, the concatamer was cloned into an entry vector (TOPO®, Invitrogen, CA). In the second step, the concatamer was transferred from the entry vector to the MEL1 reporter using the Gateway® LR Clonase™ system (Invitrogen), essentially as described by the manufacturer. Reporter constructs containing target sites for CCR5-, IPP2K-, and POU5F1/Oct34-targeted ZFNs were generated and used for normalization.

Example 2: Integration of the Reporter Construct into Yeast

[0107] The integration of the reporter construct into the 69-1B strain (S288C background; *MAT α his3 Δ 200 lys2-128 δ leu2 Δ 1*) was performed as described (Voth et al. (2001) *Nucleic Acids Res* 29:E59-59). The correct integration was confirmed by

5 Colony PCR using the following oligos: HO-L: 5'-

TATTAGGTGTGAAACCACGAAAAGT-3' (SEQ ID NO:7); 5'-

ACTGTCATTGGGAATGTCTTATGAT-3' (SEQ ID NO:8); HO-R: 5'-

attacgctcgcatcaaaatca-3' (SEQ ID NO:9); and 5'-

CATGTCTTCTCGTTAAGACTGCAT-3' (SEQ ID NO:10).

Example 3: ZFN Activity Assay

[0108] To demonstrate that cleavage at the target site of the reporter construct restores MEL1 activity, the following experiments were performed. A SSA reporter construct was engineered as described above to include a recognition site of the HO endonuclease and integrated into host cells as described above. The cells were then transfected with expression vectors encoding the HO endonuclease and the cells cultured in the presence or absence of galactose.

[0109] Results are summarized in Table 1.

20 Table 1: Cleavage at target site of the reporter construct restores MEL1 activity

Host Cell	Expression construct	Galactose	MEL1 Activity
Reporter – no HO target site	empty vector	-	-
Reporter – no HO target site	empty vector	+	-
Reporter – no HO target site	pGAL HO	-	-
Reporter – no HO target site	pGAL HO	+	-
Reporter with HO target site	empty vector	-	-
Reporter with HO target site	empty vector	+	-

Reporter with HO target site	pGAL HO	-	-/+
Reporter with HO target site	pGAL HO	+	+++

[0110] As shown in Table 1, MEL1 activity was only observed in the presence of the HO endonuclease, and of its target in the reporter locus. Furthermore, essentially no MEL1 gene activity was observed when HO expression was not induced, with the exception of very low-frequency spontaneous MEL1 restoration events. Induction of HO endonuclease expression converted essentially 100% of the cells in the sample to a MEL1 state.

Example 4: Identification of Persistently Biologically Active NME-specific ZFNs

[0111] ZFNs were designed to recognize sequences within NME and plasmids comprising sequences encoding these designed NME ZFNs were constructed essentially as described in Urnov et al. (2005) *Nature* 435(7042):646-651. The ZFNs were tested in *in vitro* assays (ELISA). Figure 3 shows the information about the NME-binding ZFNs and their DNA binding characteristics *in vitro*.

[0112] For *in vivo* screening, the coding sequences of the ZFNs were transferred to galactose inducible expression vectors using standard cloning procedures (Moehle et al. (2007) *Proc Natl Acad Sci U S A* 104:3055-3060; Mumberg et al. (1994) *Nucleic Acids Res* 22:5767-5768; Urnov et al. (2005) *Nature* 435:646-651).

[0113] The NME1 cDNA (RZPD IRAUp969A091D6) target site was subcloned into the reporter construct and the resulting reporter construct was integrated into the genome of the yeast strain as described in Example 2.

A. Reporter Activity

[0114] Expression constructs encoding NME1-targeted ZFN pairs were transformed into the reporter strain in deep well blocks as described in Gietz and Woods, (2006) *Methods Mol Biol* 313:107-120. In order to eliminate the cumbersome manipulation of Petri dishes, pools of transformants were selected in liquid media. Briefly, the cells were resuspended in 1 ml of SC His-Leu-media and incubated for 48

hours at 30°C. To further enrich for transformants, a 1:10 dilution of the pool was incubated in fresh media for another 24 hours.

[0115] To de-repress the GAL1 promoter, the pools of transformants were diluted 1:10 into 1 mL of SC His-Leu- media containing 2% raffinose as a source of carbon and incubated O/N at 30°C. ZFN expression was induced by diluting the raffinose cultures 1:10 into 1ml of SC His-Leu- media containing 2% galactose. Cells were then incubated for various amount of time, typically from 2 to 6 hours, before addition of 2% glucose to stop expression. Cells were then incubated overnight to allow for DSB repair and reporter gene expression.

[0116] In order to normalize the reporter signal to the amount of cells in the culture, spectrophotometric readings of the well blocks were taken at 600 nm. The deep well block was then centrifuged at 3000g for 5 minutes to pellet yeast cells and 10 µl of the media is assayed for Mel1 activity as described in Chen et al. (2004) *Anal Biochem* 335:253-259 and Ryan et al. (1998) *Mol Cell Biol* 18:1774-1782.

[0117] Results of reporter gene expression are shown in FIG. 4A. The ZFN pairs are indicated below the bars. For each pair, the left-most bar shows Mel1 activity prior to introduction of the indicated ZFN pairs, the second bar from the left shows Mel1 activity 2 hours after the indicated ZFN pairs are introduced into the cells; the third bar from the left shows Mel1 activity 4 hours after the indicated ZFN pairs are introduced into the cells; and the right-most bar shows Mel1 activity 6 hours after the indicated ZFN pairs are introduced into the cells. As shown, there is no Mel1 activity in the absence of an active ZFN pair. Furthermore, although all ZFN pairs were active to some degree, the assay provided a ranking of the most active pairs.

B. Toxicity

[0118] To evaluate toxicity of the various ZFNs, ZFN expression was induced by diluting the raffinose cultures 1:100 into 1ml of SC His-Leu-media containing 2% galactose. Cells were then incubated for various amount of time, typically from 24 to 30 hours. Growth of the populations were then determined by spectrophotometry reading at 600 nm.

[0119] Yeast cell growth in the presence of ZFNs is shown in Fig. 4B. The ZFN pairs are indicated below the bars. For each pair, the left-most bar shows OD600 prior to introduction of the indicated ZFN pairs, the second bar from the left

shows OD600 readings 23 hours after the indicated ZFN pairs are introduced into the cells; the third bar from the left shows OD₆₀₀ 27 hours after the indicated ZFN pairs are introduced into the cells; and the right-most bar shows OD600 30 hours after the indicated ZFN pairs are introduced into the cells. As shown, yeast cells grow

5 normally in the presence of certain ZFN pairs.

C. Yeast Screen Predicts in vivo Activity

[0120] Various active ZFN pairs identified in the reporter activity assay (Fig. 4A) were then transformed into human K562 and tested for their ability to induce
10 mutations in the NME 1 locus, as described in Miller et al. (2007) *Nat Biotechnol.* 25(7):778-85.

[0121] As shown in Fig. 5, the most active ZFN pair against NME1 in a human cultured cell line is 13674-13677. Furthermore, this ZFN pair was identified above as one of the most active proteins (Fig. 4A) that did not affect yeast cell growth
15 (Fig. 4B).

[0122] One of the major problems associated with gene modification in human cell lines is the gradual loss of modified cells over time. With ZFNs, for example, after nine days in culture, the percent of gene modified cells often drops off, perhaps due in part to the toxicity associated with overexpression of non-specific
20 ZFNs. See, in Miller et al. (2007) *Nat Biotechnol.* 25(7):778-85. Accordingly, to further confirm that the ZFN pair identified from the yeast reporter and toxicity screens remained nontoxic over time, the percent modification of human chromosomes in K562 cells was measured 2 days and 9 days after introduction of the ZFN pair 13674-13677.

25 [0123] As shown in Fig. 6, the percent modification of the NME1 locus observed two days (D2) after transfection of the 13674-13677 ZFN pair was maintained after nine days (D9) of culture.

[0124] These results demonstrate the yeast system described herein accurately predicts *in vivo* ZFN activity and, in addition, also accurately predicts which ZFNs
30 give a persistent signal over time.

Example 5: Identification of Persistently Biologically Active PD1-specific ZFNs

[0125] ZFNs were assembled against the human PD1 gene and were tested by ELISA and CEL1 assays as described in Miller et al. (2007) *Nat. Biotechnol.* 25:778-

785 and U.S. Patent Publication No. 20050064474 and International Patent Publication WO2005/014791.

[0126] From this initial *in vitro* screen, two lead ZFN pairs were identified and submitted for elaboration in order to try to improve their efficiency. These pairs target exons 1 and 5 of this gene, respectively. The elaborated (improved) proteins were retested in a time-course experiment, essentially as described in Example 4 above. The results are summarized in Table 2 below.

Table 2: PD1 NHEJ

Target	ZFN pair	% NHEJ		
		Day 3	Day 7	Day 9
exon 1	12942/12946	8	7	5
exon 1	12942/12947	10	6	6
exon 5	12934/12971	11	6	1.5
exon 5	12934/12972	11	7.5	2

[0127] As shown in Table 2, treatment of cells with ZFNs against exon 5 causes the loss of a greater proportion of genome-edited cells from the population, while the genome-editing signal in cells treated with ZFNs designed against exon 1 is much more stable.

[0128] These ZFNs were also tested in the yeast system for activity and for toxicity. Briefly, the PD1 cDNA (NM_005018) was subcloned into the SSA reporter construct and the assays performed as described in Example 3.

[0129] As shown in Fig. 7, the *in vivo* assay system clearly confirmed the activity of the PD1 ZFNs (Fig. 7A) and also determined the toxicity for all ZFNs targeting exon 5 (Fig. 7B). These results correlate with the loss of signal detected in human cells. Furthermore, the ZFNs designed against exon 1 show no impairment of yeast growth and maintain the signal in human cells.

Example 6: Identification of Biologically Active Zebrafish ZFNs

[0130] Panels of ZFNs were assembled against the zebrafish SLC24A5 (“golden”) and *notail* (“NTL”) genes and screened for activity in a yeast host cells comprising an SSA reporter construct with the appropriate target sequence. See, also,

U.S. Serial No. _____, "Genomic Editing in Zebrafish Using Zinc Finger Nucleases," filed even date herewith.

[0131] Results of yeast activity and toxicity screens are shown in Fig. 9A (*golden* ZFN activity); Fig. 9B (*golden* ZFN toxicity); Fig. 10A (*notail* ZFN activity); and Fig. 10B (*notail* ZFN toxicity).

[0132] ZFNs identified by the yeast screen were then injected into zebrafish embryos. Only embryos expressing the *golden*-ZFNs exhibited somatic mosaicism for pigmentation (Fig. 11). Similarly, only embryos expressing the *NTL*-ZFNs exhibited the forked-tail phenotype (Fig. 12).

[0133] Thus, the *in vivo* assay system described herein identified ZFNs that are biologically active in zebrafish.

Example 7: Identification of active ZFN variants using a selection-based screen

[0134] Active ZFN variants were selected using a positive and negative selection reporter yeast strain as follows. The reporter construct (Fig. 8) was generated and contained a homodimer recognition site for a well characterized ZFN (8266). See, Table 1 of U.S. Patent Publication No. 20080159996 for recognition helix and target sequence of 8266. The construct also contained a counterselectable gene (URA3) between the interrupted MEL1 gene as described in Example 4. In yeast cells containing this reporter, cells containing an inactive ZFN will be killed in the presence of 5-FOA permitting the selective expansion of cells containing active ZFN variants in the population. Similarly, cells containing an active ZFN cannot grow in the absence of uracil.

[0135] ZFNs containing five different linker sequences (different in length and/or amino acid residues) between the ZFP and the FokI cleavage domain were prepared and tested in the yeast screen assay described above. ZFNs with the wild-type linker sequence (8266) were active. In addition, two moderately active and two inactivate variants were identified. A pool of ZFNs containing 8266 as well as the two moderately active and two inactive variants was prepared and tested in a yeast counterselection assay as follows.

[0136] The ZFN variants were mixed in equimolar ratios and the reporter strains was transformed as described in Example 4. ZFN induction was performed as

in Example 4 for 6 and 24 hours. After a recovery period of 16 hours, the cells were diluted 100 fold in media containing the counterselecting agent 5-FOA for 20 hours. As a verification that the active ZFNs had cleaved the URA3 gene, the cells were similarly diluted into media lacking added uracil. As can be seen from the right
5 portion of Figure 13, the cells containing active ZFN (SBS 8266) were not able to grow in the absence of uracil. As shown in Figure 13, yeast growth is observed only when active ZFN were used to transform the reporter strain.

[0137] The cells were then harvested, total DNA was extracted from the yeast population and individual plasmids were obtained by retransforming into *E. coli*.

10 Twelve colonies were picked at random and sequenced. Of the 12, 9 sequences matched the wild-type linker, 3 sequences corresponded to the moderately active variants, while no sequences corresponded to the inactive variants. Thus, it was possible to specifically enrich the most active variants within a population of ZFN
15 variants.

[0138] These data validate the yeast system for the identification of biologically active ZFN pairs from a set of biochemically active variants and also demonstrate the use of the toxicity assay for the determination of their relative
20 specificity.

[0139] All patents, patent applications and publications mentioned herein are hereby incorporated by reference in their entirety.

[0140] Although disclosure has been provided in some detail by way of
25 illustration and example for the purposes of clarity of understanding, it will be apparent to those skilled in the art that various changes and modifications can be practiced without departing from the spirit or scope of the disclosure. Accordingly, the foregoing descriptions and examples should not be construed as limiting.

CLAIMS

What is claimed is:

- 5 1. A reporter construct for detecting double-stranded cleavage of a target sequence by one or more nucleases, the reporter construct comprising overlapping and non-functional sequences of a reporter gene separated by a target sequence recognized by the nuclease.
- 10 2. The reporter construct of claim 1, wherein the reporter gene is operably linked to a promoter sequence.
3. The reporter construct of claim 2, wherein the promoter is constitutive.
- 15 4. The reporter construct of claim 2, wherein the promoter is an inducible promoter.
5. The reporter construct of any of claims 1 to 4, wherein the reporter gene encodes an enzyme.
- 20 6. The reporter construct of claim 5, further comprising a sequence encoding a selectable marker.
7. The reporter construct of any of claims 1 to 6, further comprising regions
25 of homology flanking the discontinuous reporter gene sequences.
8. A host cell comprising a reporter construct according to any of claims 1 to 7.
- 30 9. The host cell of claim 8, wherein the cell is a eukaryotic cell.
10. The host cell of claim 9, wherein the cell is a yeast cell.

11. The host cell of any of claims 8 to 10, wherein the reporter construct is transiently expressed in the host cell.

12. The host cell of any of claims 8 to 10, wherein the reporter construct is
5 stably integrated into the host cell.

13. A method of identifying a nuclease that induces cleavage at a specific target site, the method comprising the steps of:

introducing one or more expression constructs that expresses the nuclease into
10 a host cell according to any of claims 8 to 12, wherein the reporter construct comprises a target sequence recognized by the nuclease;

incubating the cells under conditions such that the nuclease is expressed; and
measuring the levels of reporter gene expression in the cells, wherein
increased levels of reporter gene expression are correlated with increased nuclease-
15 induced cleavage of the target sequence.

14. A method of ranking a panel of nucleases for their activity in inducing cleavage at a specific target site, the method comprising the steps of:

introducing one or more expression constructs encoding nucleases of the panel
20 into separate host cells according to any of claims 8-12, wherein the reporter construct in the host cell comprises a target sequence recognized by the nuclease;

incubating the cells under conditions such that the nucleases are expressed;
measuring the levels of reporter gene expression in the cells; and
ranking the nucleases according to levels of reporter gene activity induced in
25 the host cell.

15. A methods of predicting the *in vivo* cleavage activity of a nuclease, the method comprising the steps of:

introducing an expression construct encoding the nuclease into a host cell
30 according to any of claims 8 to 12, wherein the reporter construct comprises a target sequence recognized by the nuclease;

incubating the cells under conditions such that the nuclease is expressed; and
measuring the levels of reporter gene expression in the cells;

wherein higher levels or reporter gene expression are predictive of a nuclease that will be active *in vivo*.

16. A method of determining toxic effects on a host cell caused by a nuclease,
5 the method comprising the steps of:
introducing an expression construct encoding the nuclease into a host cell;
incubating the cells under conditions such that the nuclease is expressed;
culturing the cells over a period of time; and
measuring the growth of cells in culture at various time intervals, wherein
10 diminished growth is correlated with increased toxic effects of the nuclease.

17. A method of selecting a biologically active nuclease, the method
comprising the steps of:
identifying a nuclease cleaves at a selected target site according to claim 13;
15 and
introducing an expression construct encoding the nuclease into a host cell;
incubating the cells under conditions such that the nuclease is expressed;
culturing the cells over a period of time; and
measuring the growth of cells in culture at various time intervals, wherein
20 biologically active nuclease(s) exhibiting cleavage activity and low toxicity are
selected.

18. The method of claim 16 or 17, wherein the growth of the cells is
determined by spectrophotometry.

25

19. The method of any of claims 16 to 18, wherein the host cell comprises a
reporter construct according to any of claims 1 to 7.

20. The method of any of claims 13 to 18, the nuclease comprises a homing
30 endonuclease.

21. The method of any of claims 13 to 18, wherein the nuclease comprises an
engineered zinc finger nuclease or pair of engineered zinc finger nucleases.

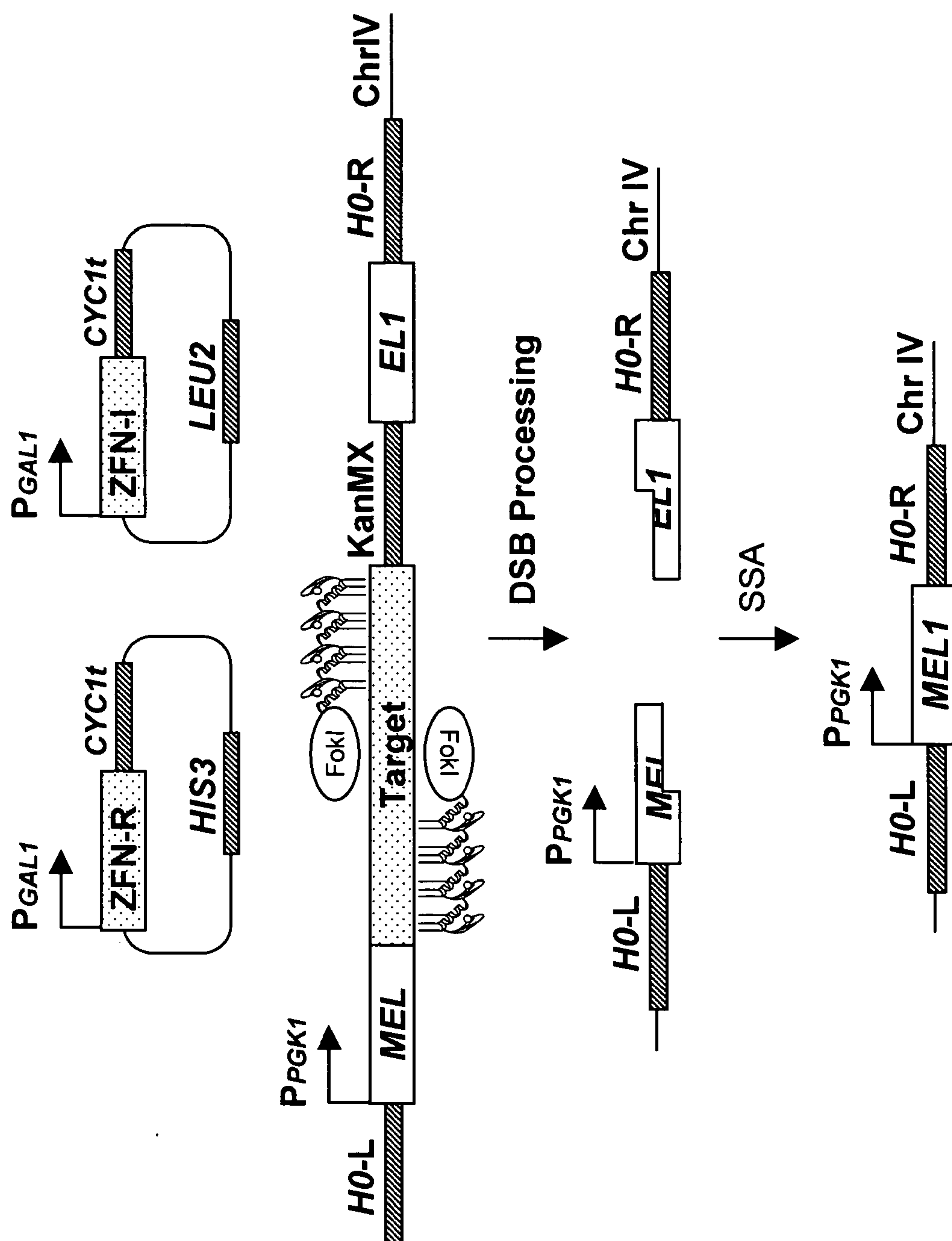


Fig. 1

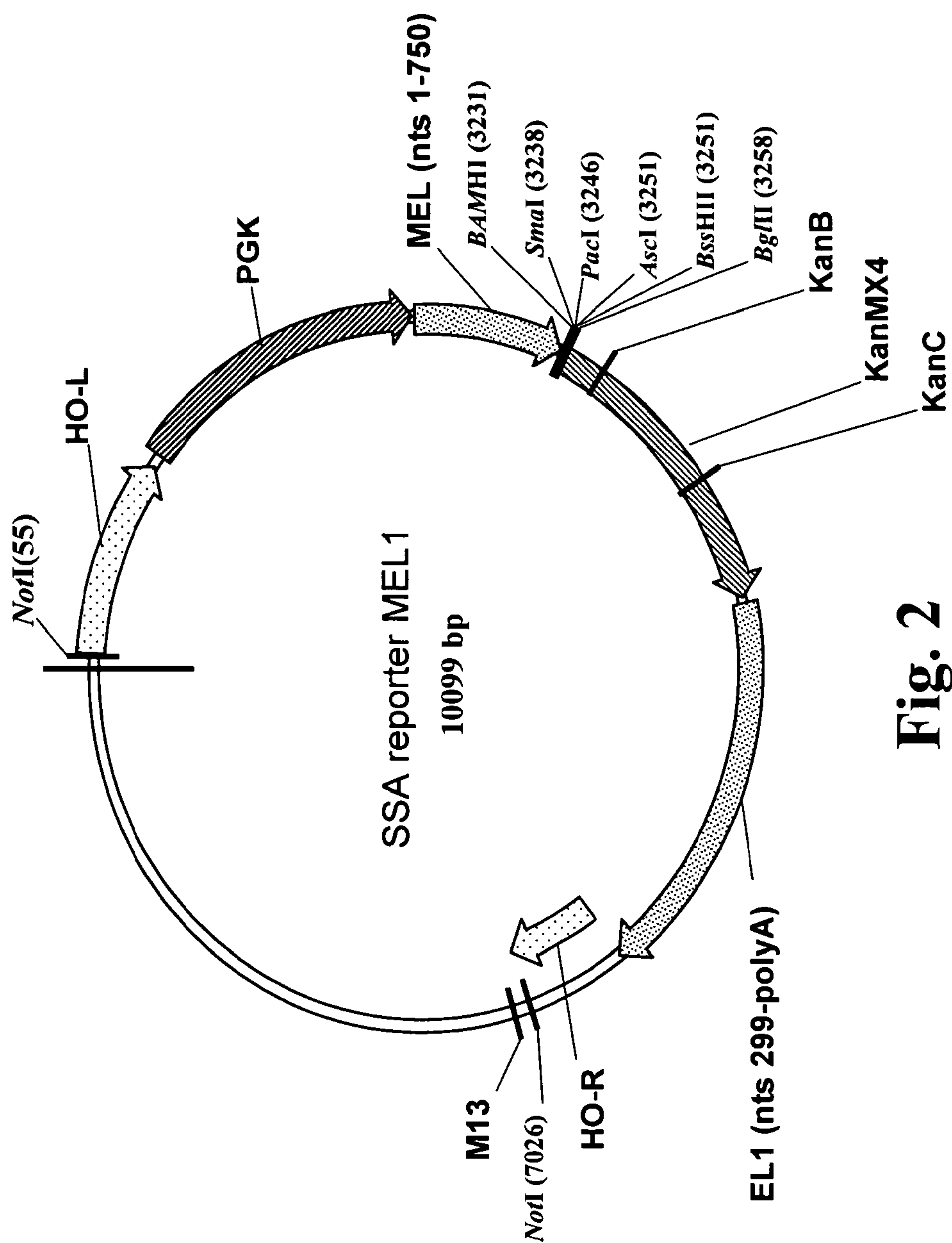


Fig. 2

SBS#	Name	Code	Site	Score	ELISA Data			Normalization		
					Exp 1	Exp 2	Avg	5475	58-59	Parental
13636	NME1-3097a1	5-E6/2-J13	ccATGGCCcAcACTGTGagcgtaccctcat	6.18	514	470	492	0.09	0.09	
13637	NME1-3091a1	5-L9.5-H5	tcTGGGGATAGGCAGgtagaggaaatgaa	10.76	3,025	3,207	3,116	0.58	0.60	
13638	NME1-3091b1	5-L9.5-H5	tcTGGGGATAGGCAGgtagaggaaatgaa	10.76	5,072	5,916	5,494	1.03	1.06	
13639	NME1-3091c1	5-L9/2-B9	tcTGGGGATAGGCAGgtagaggaaatgaa	7.44	2,218	3,343	2,781	0.52	0.54	
13640	NME1-3091d1	5-L9/2-N16	tcTGGGGATAGGCAGgtagaggaaatgaa	5.59	2,100	3,069	2,585	0.48	0.50	
13641	NME1-3102a1	3-H12-L17	gcCAACTGTGAGCGtaccctcattgcga	10.05	803	1,215	1,009	0.19	0.19	
13642	NME1-3102b1	3-H12-L17	gcCAACTGTGAGCGtaccctcattgcga	10.05	5,186	4,078	4,632	0.87	0.89	
13643	NME1-3102c1	6-J20.2-L17	gcCAACTGTGAGCGtaccctcattgcga	8.26	1,158	1,155	1,157	0.21	0.22	
13644	NME1-3102d1	6-J20.2-L17	gcCAACTGTGAGCGtaccctcattgcga	8.26	2,685	3,207	2,946	0.55	0.57	
13645	NME1-r3096a1	1-M3.4-K4	atGGTTCTGGGGATagcaggtagaggga	6.58	3,354	3,132	3,243	0.61	0.63	
13646	NME1-r3096b1	1-M3.4-K4	atGGTTCTGGGGATagcaggtagaggga	6.58	4,731	5,723	5,227	0.98	1.01	
13647	NME1-3131a1	2-K18.5-E11	caAACCCAGATGGGGtcacagcgggcttt	9.73	3,815	4,635	4,225	0.79	0.81	
13648	NME1-3131b1	2-K18.5-E11	caAACCCAGATGGGGtcacagcgggcttt	9.73	3,554	3,550	3,552	0.67	0.68	
13649	NME1-3131c1	6-B17.5-E11	caAACCCAGATGGGGtcacagcgggcttt	8.63	2,602	2,787	2,695	0.50	0.52	
13650	NME1-3131d1	2-K18.4-E23	caAACCCAGATGGGGtcacagcgggcttt	8.23	3,772	3,976	3,874	0.73	0.75	
13651	NME1-3125a1	6-E4.5-A13	tcGCAATGAAGGTAcgctcacagttggc	6.14	3,476	4,019	3,748	0.70	0.72	
13652	NME1-3125b1	6-E4.5-A13	tcGCAATGAAGGTAcgctcacagttggc	6.14	3,567	5,347	4,457	0.84	0.86	
13653	NME1-3218a1	2-H18.2-L20	caTGCAAGTAAGTGgacttcattgttc	6.34	281	358	320	0.06	0.06	
13654	NME1-3218b1	2-H18.2-L20	caTGCAAGTAAGTGgacttcattgttc	6.34	4,551	5,544	5,047	0.95	0.97	
13655	NME1-3218c1	2-H18.4-O24	caTGCAAGTAAGTGgacttcattgttc	5.25	455	620	537	0.10	0.10	
13656	NME1-r3211a1	5-I2/2-K17	ttTCAGACcAACAAAGcggaatccttc	5.29	3,439	5,514	4,476	0.84	0.86	
13657	NME1-7521a1	5-E15.2-H11	ccATGGTGAGTGTGcctgtgtggtatac	8.63	301	512	406	0.07	0.08	
13658	NME1-7521b1	5-E15.2-H11	ccATGGTGAGTGTGcctgtgtggtatac	8.63	859	1,408	1,134	0.21	0.22	
13659	NME1-7521c1	4-E24.2-H11	ccATGGTGAGTGTGcctgtgtggtatac	8.63	34	78	56	0.01	0.01	
13660	NME1-r7515a1	3-N4.2-J7	aaCTACCGGCCCTGagtgcatgtatttc	8.01	1,000	1,931	1,466	0.27	0.28	
13661	NME1-r7515b1	3-N4.2-J7	aaCTACCGGCCCTGagtgcatgtatttc	8.01	2,911	3,564	3,237	0.61	0.62	
13662	NME1-r7515c1	3-N4.6-F8	aaCTACCGGCCCTGagtgcatgtatttc	6.07	300	599	449	0.08	0.09	
13663	NME1-8676a1	5-L2.3-P2	ccTGGGACCATCCGtggaactctgca	9	1,580	2,456	2,018	0.38	0.39	
13664	NME1-8676b1	5-L2.3-P2	ccTGGGACCATCCGtggaactctgca	9	2,562	3,008	2,785	0.52	0.54	
13665	NME1-r8670a1	3-K17.2-F18	ttGGAGTCTGCAGGgttggctccccga	5.56	2,358	3,172	2,765	0.52	0.53	
13666	NME1-r8670b1	3-K17.2-F18	ttGGAGTCTGCAGGgttggctccccga	5.56	3,504	5,659	4,582	0.86	0.88	
13667	NME1-9178a1	2-J23.4-F15	taTACATGGCAGTGattctgtggagagt	5.97	3,819	3,563	3,691	0.69	0.71	
13668	NME1-9178b1	2-J23.4-F15	taTACATGGCAGTGattctgtggagagt	5.97	3,930	5,736	4,833	0.91	0.93	
13669	NME1-9178c1	6-N16.4-F15	taTACATGGCAGTGattctgtggagagt	5.33	3,762	6,366	5,064	0.95	0.98	
13670	NME1-9171a1	3-O6.4-O3	gtTCCTGGGTGGAGaaagagatagtcac	9.25	5,123	6,320	5,721	1.07	1.10	
13671	NME1-r9171b1	3-O6.4-O3	gtTCCTGGGTGGAGaaagagatagtcac	9.25	3,377	4,818	4,097	0.77	0.79	
13672	NME1-r9171c1	3-O6.4-O17	gtTCCTGGGTGGAGaaagagatagtcac	9.25	5,637	6,869	6,253	1.17	1.21	
13673	NME1-r9171d1	6-P23.4-O3	gtTCCTGGGTGGAGaaagagatagtcac	9.25	4,666	5,478	5,072	0.95	0.98	
13674	NME1-9289a1	2-F8.4-E2	acAGGAGGGCAGACcaccattgcttcca	5.11	4,088	6,156	5,122	0.96	0.99	
13675	NME1-9289b1	2-F8.4-E2	acAGGAGGGCAGACcaccattgcttcca	5.11	3,029	4,740	3,885	0.73	0.75	
13676	NME1-r9283a1	3-I4.2-M22	atTCATAGATCCAGttctgacacagct	5.68	36	64	50	0.00	0.01	
13677	NME1-r9283b1	3-I4.2-M22	atTCATAGATCCAGttctgacacagct	5.68	2,327	2,978	2,653	0.50	0.51	
13678	NME1-r9283c1	6-O21.2-M22	atTCATAGATCCAGttctgacacagct	5.55	34	44	39	0.00	0.01	
13679	NME1-r9283d1	3-G4.2-M22	atTCATAGATCCAGttctgacacagct	5.51	1,267	1,899	1,583	0.29	0.31	

Fig. 3

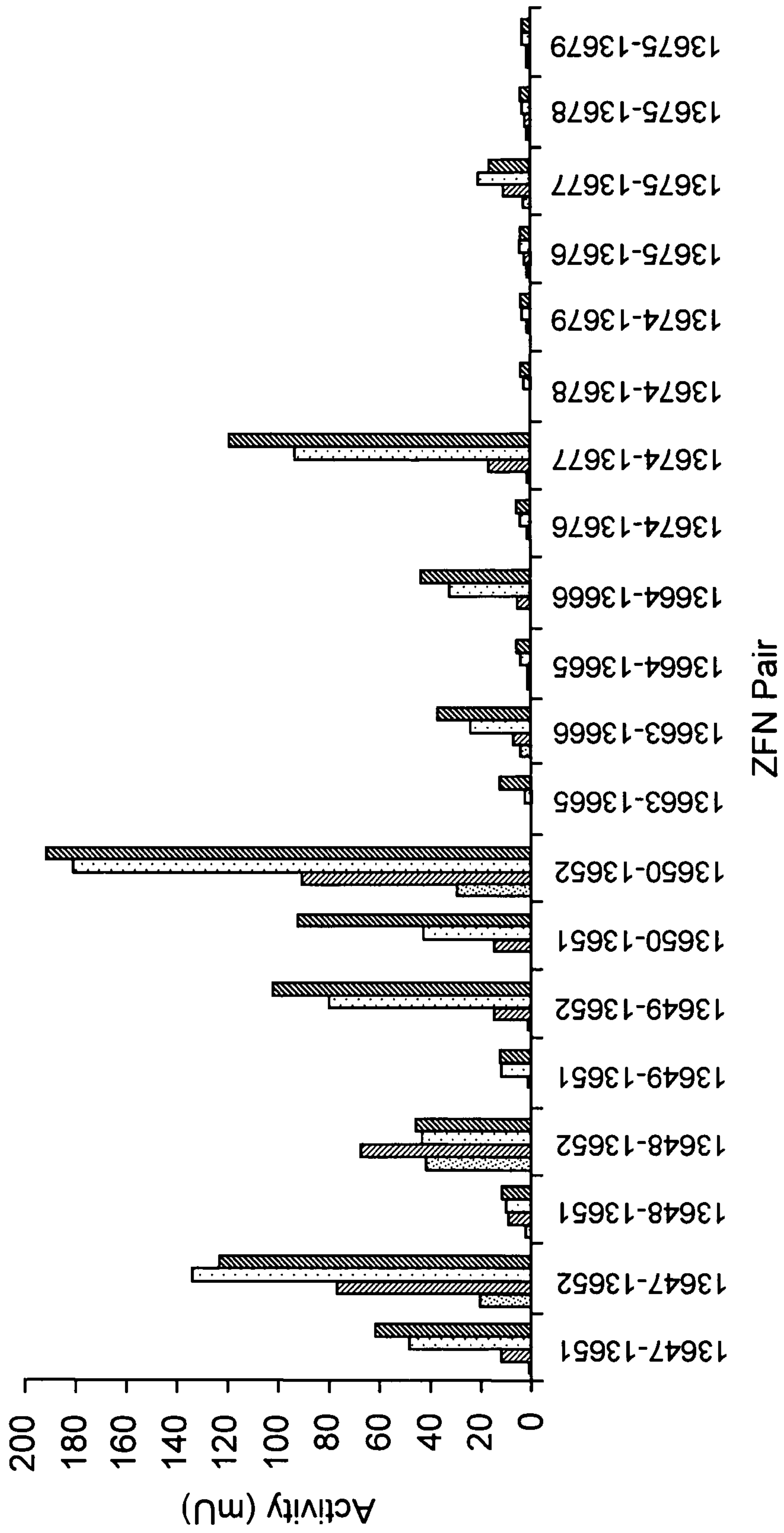
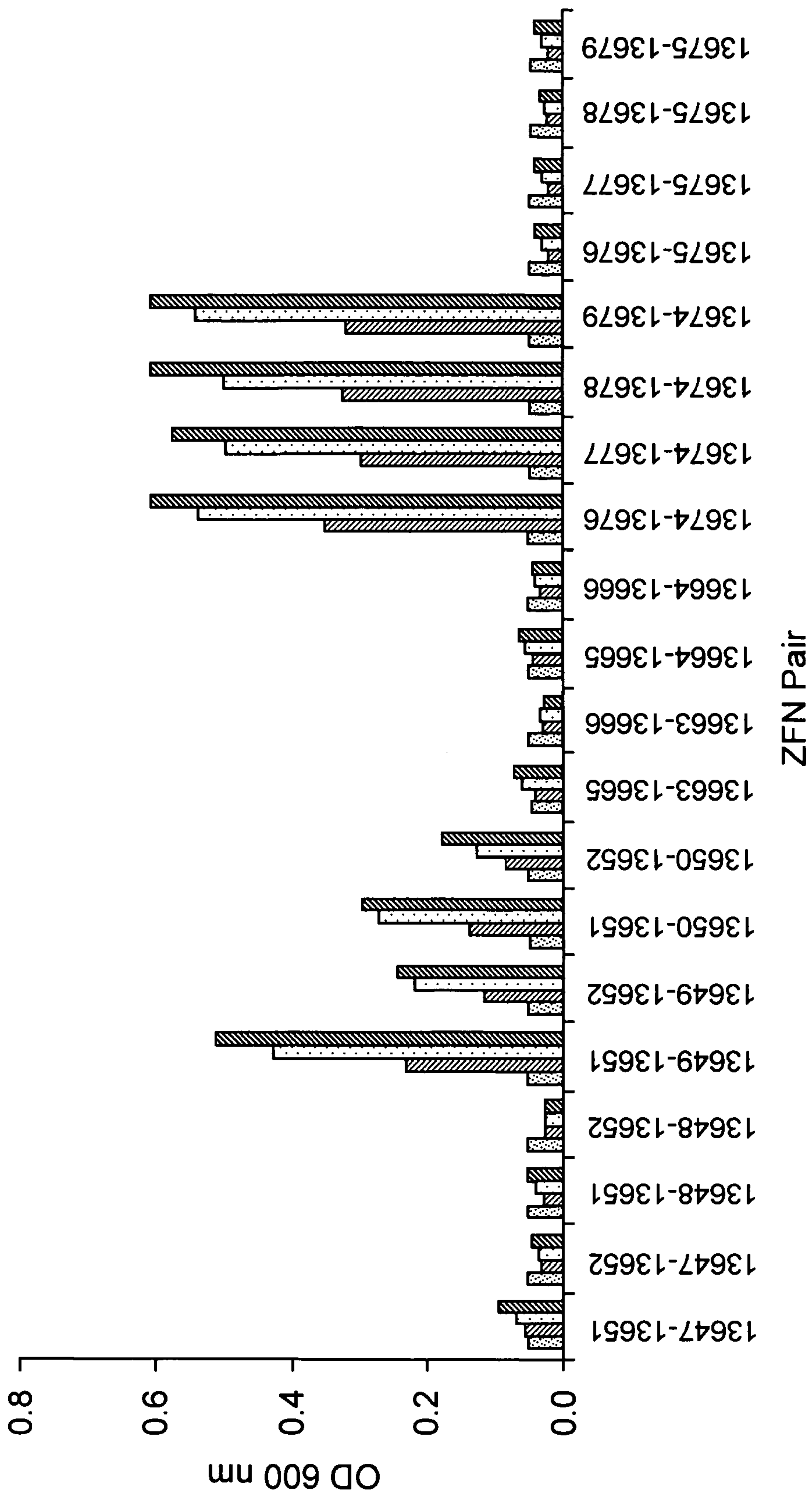


Fig. 4A

**Fig. 4B**

1	GFP
2	13647 & 13652
3	13649 & 13652
4	13650 & 13651
5	13650 & 13652
6	GFP
7	13674 & 13677
8	GFP
9	AAVS1 Positive control

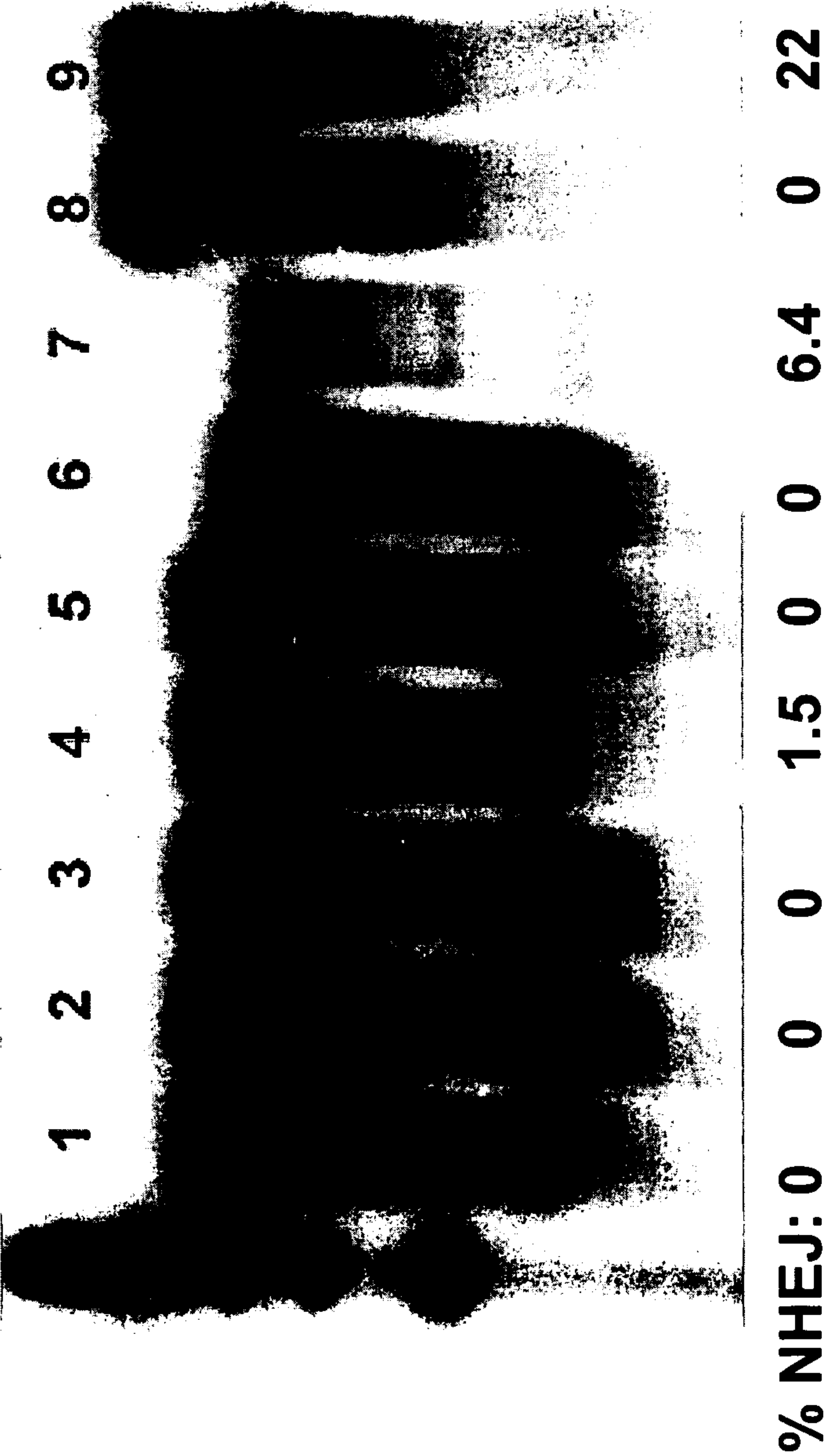


Fig. 5

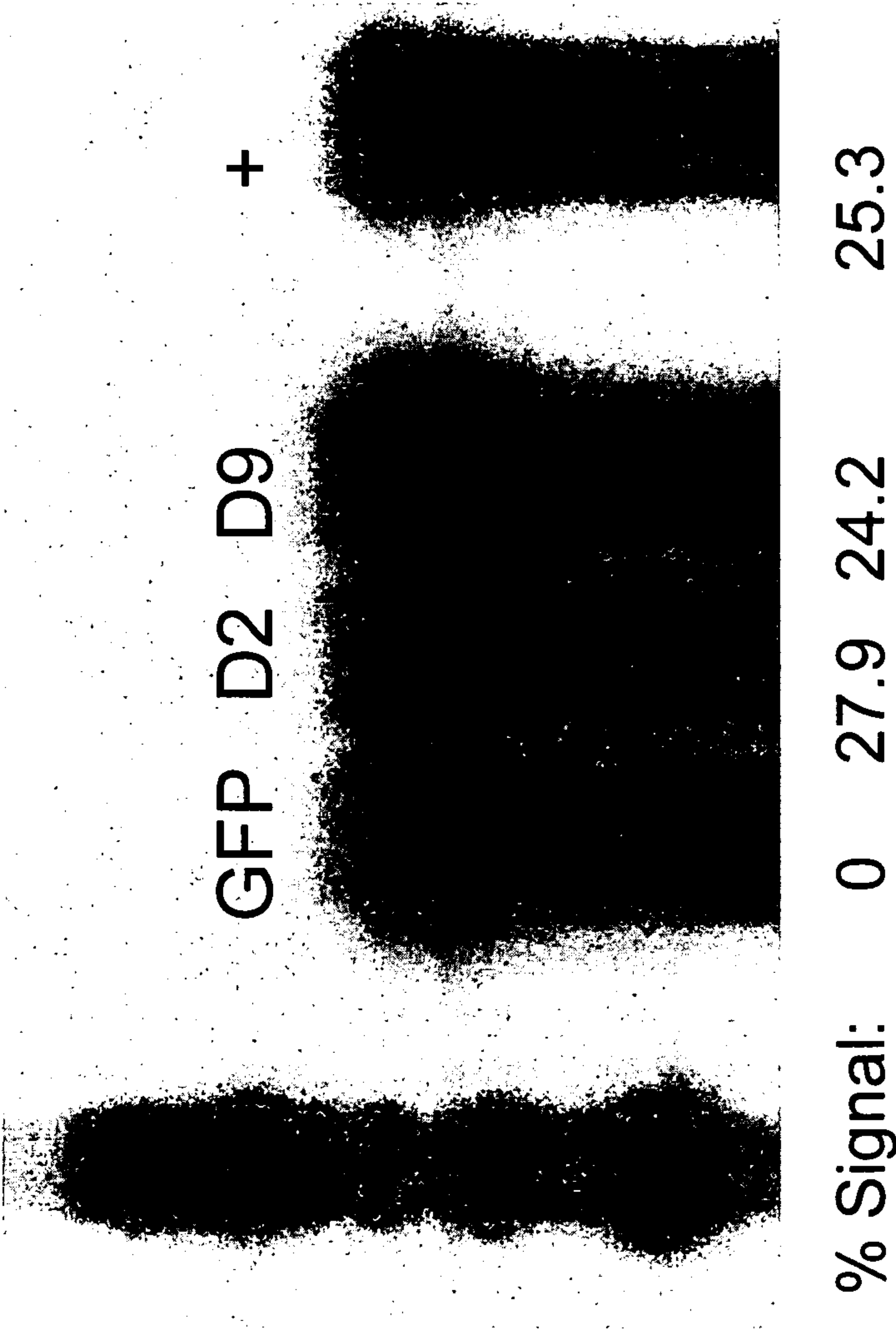


Fig. 6

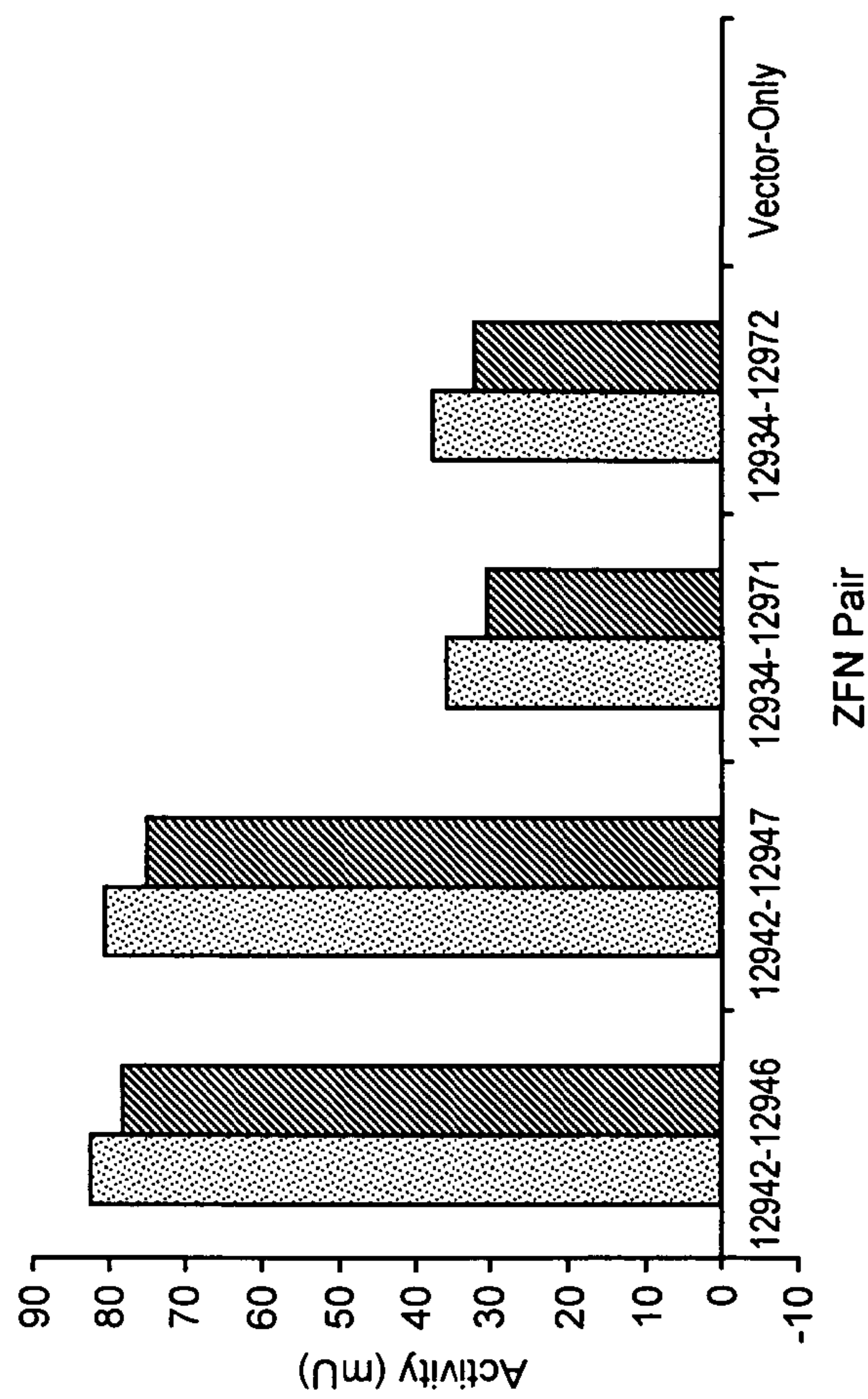


Fig. 7A

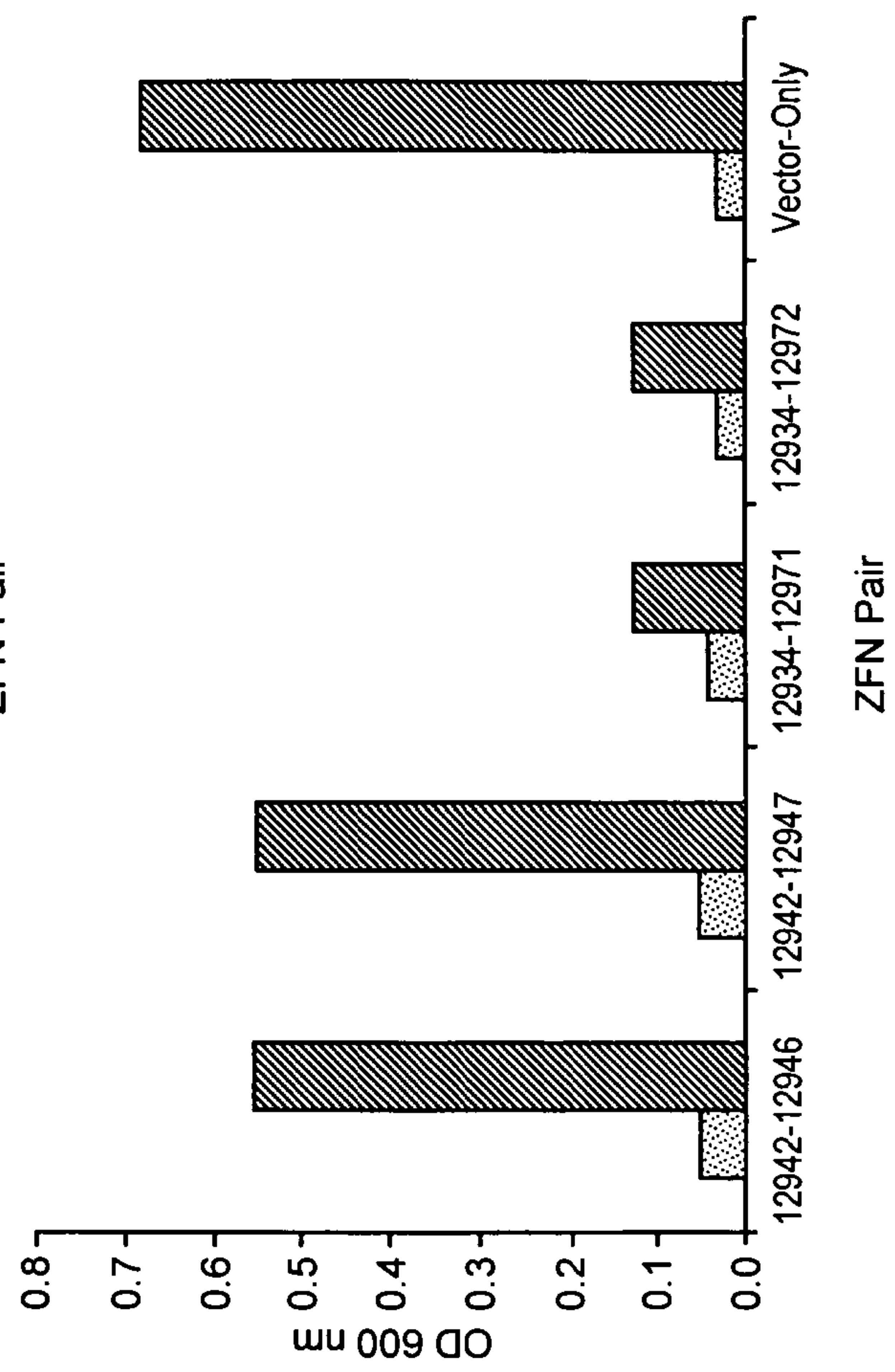


Fig. 7B

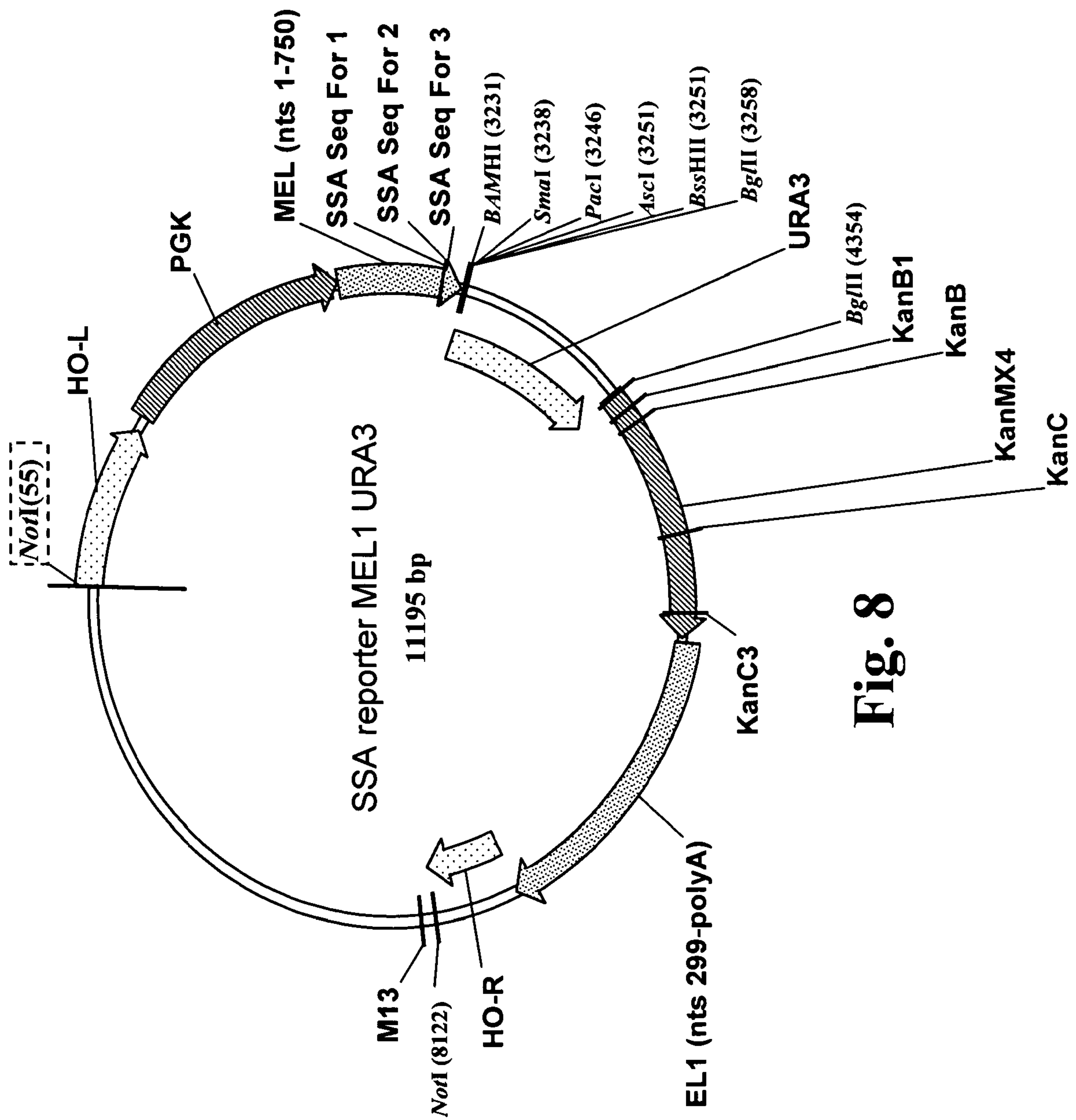
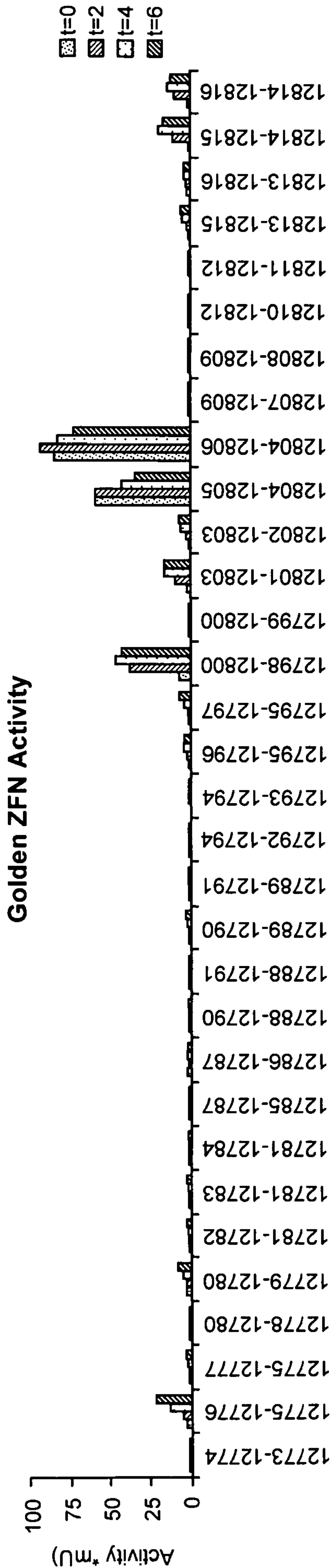
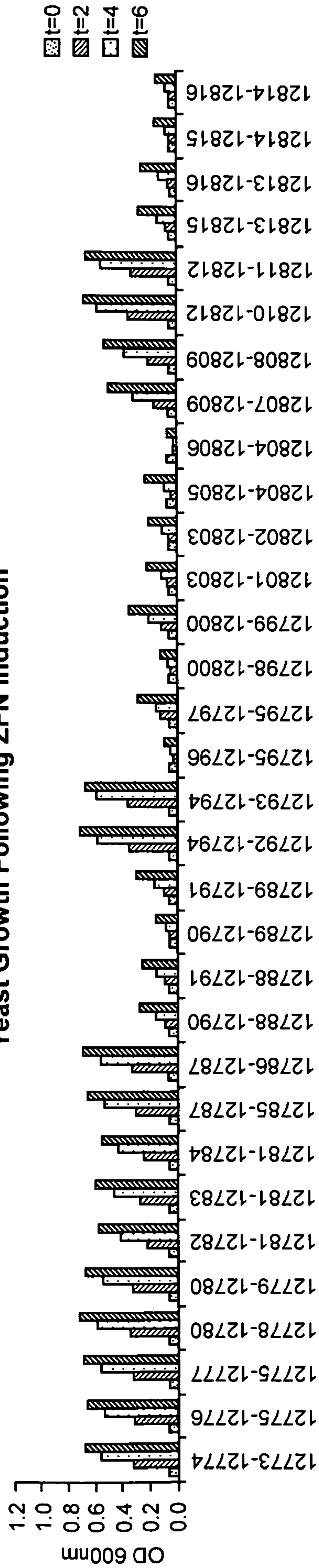


Fig. 8



Yeast Growth Following ZFN Induction



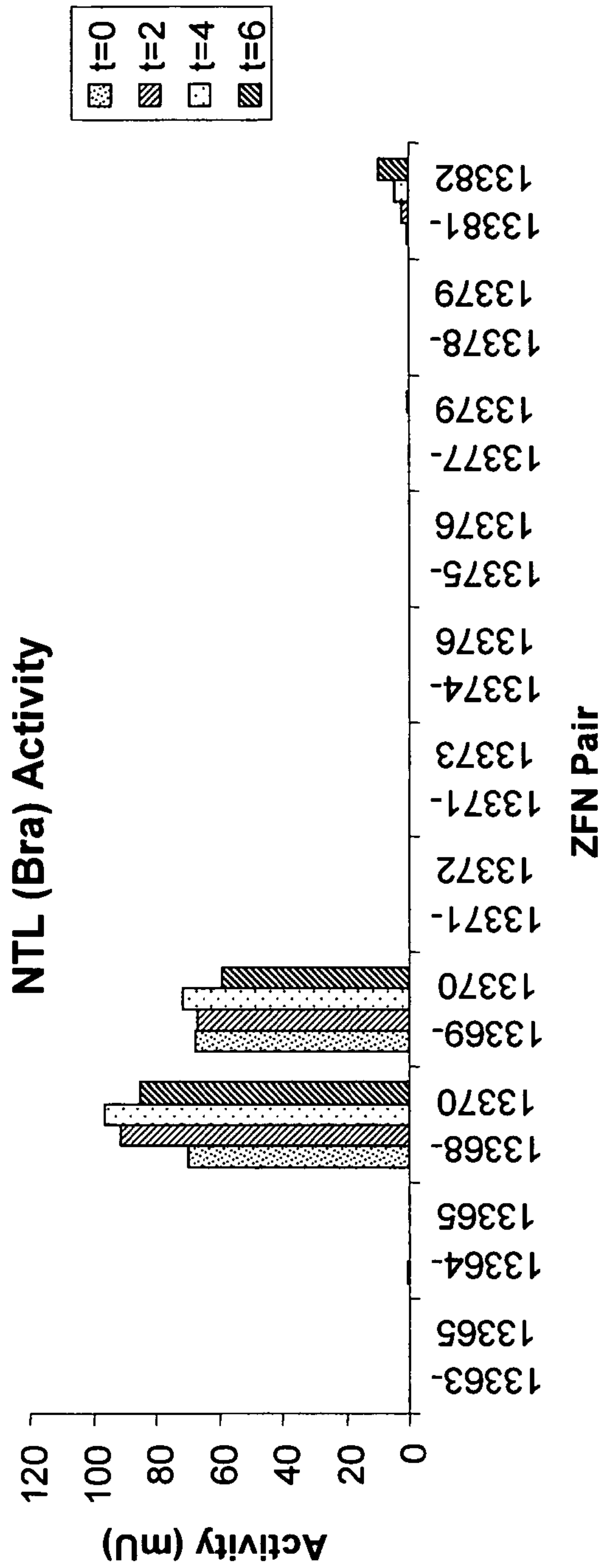


Fig. 10A

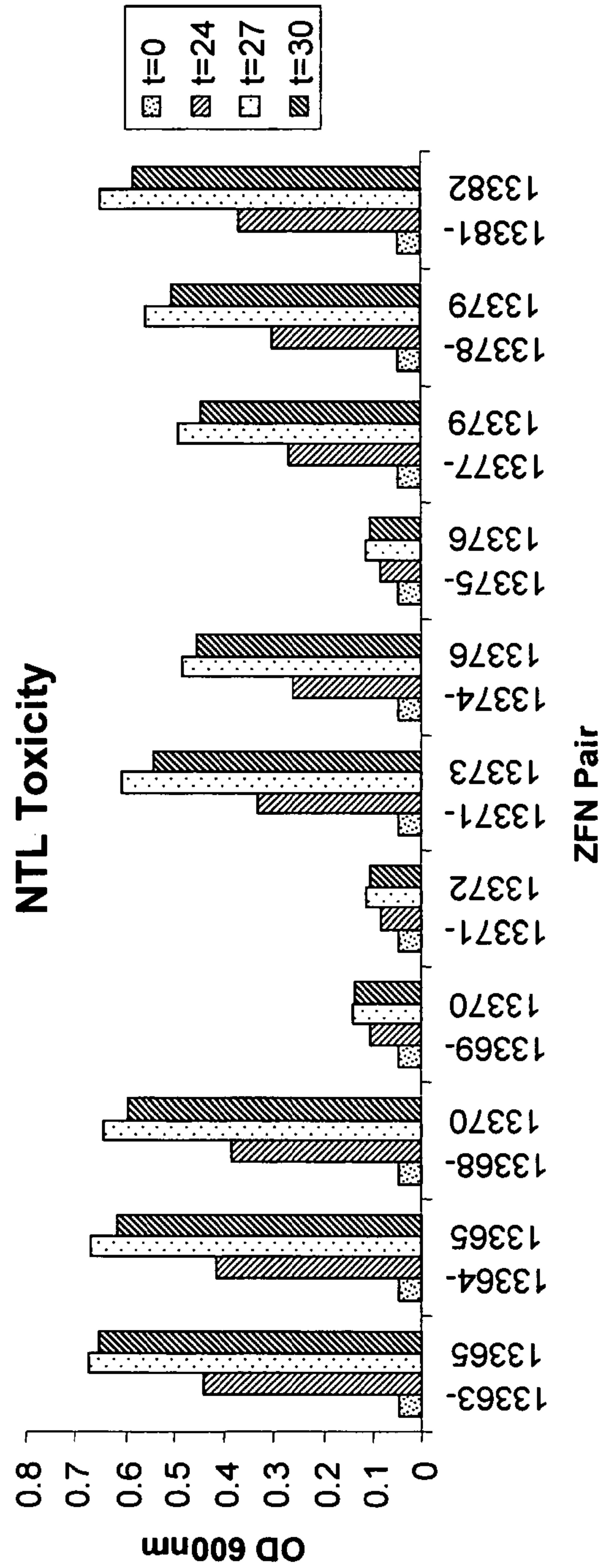
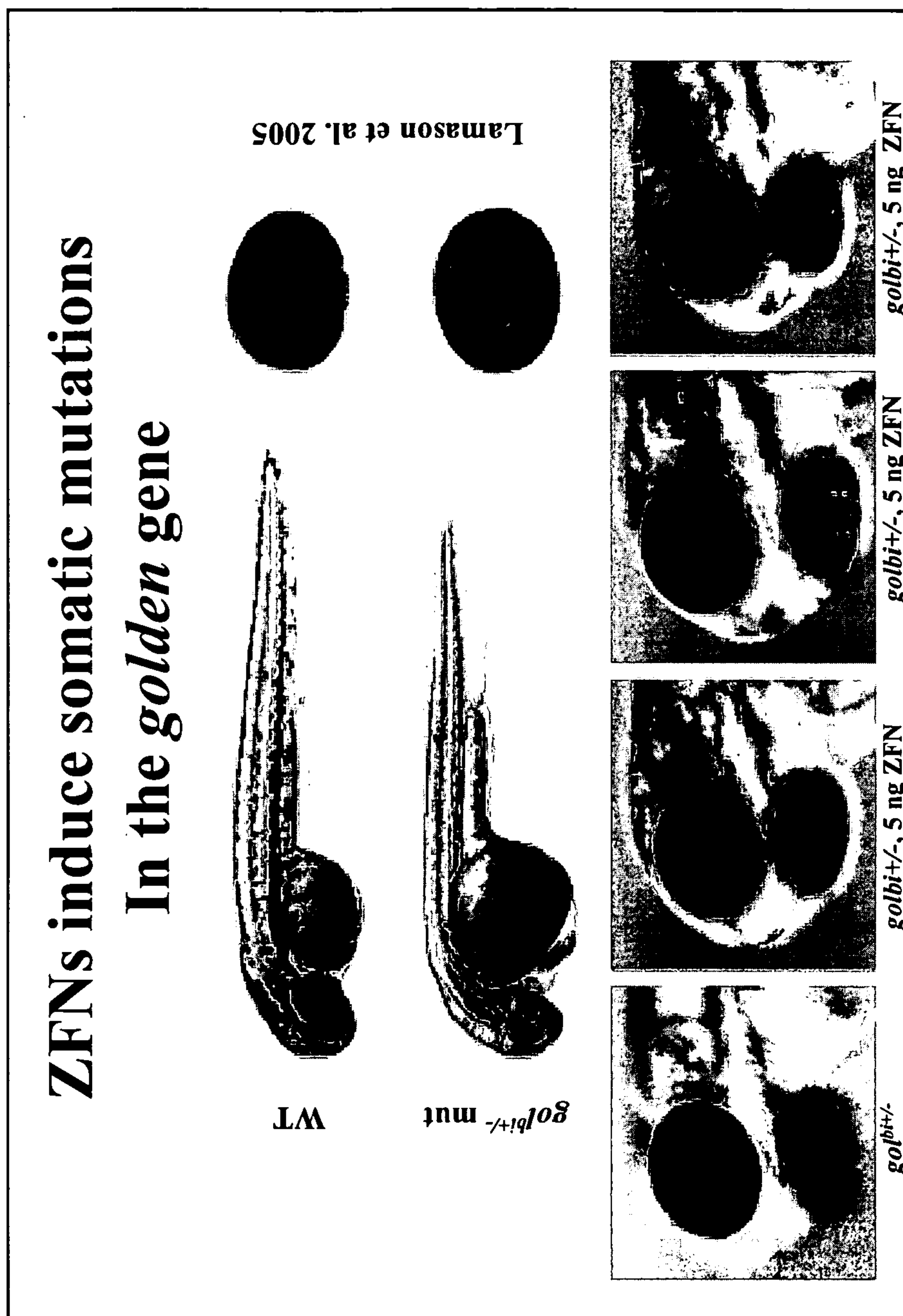


Fig. 10B

**Fig. 11**

ZFNs Induce somatic mutations In the *no tail/Brachyury* gene

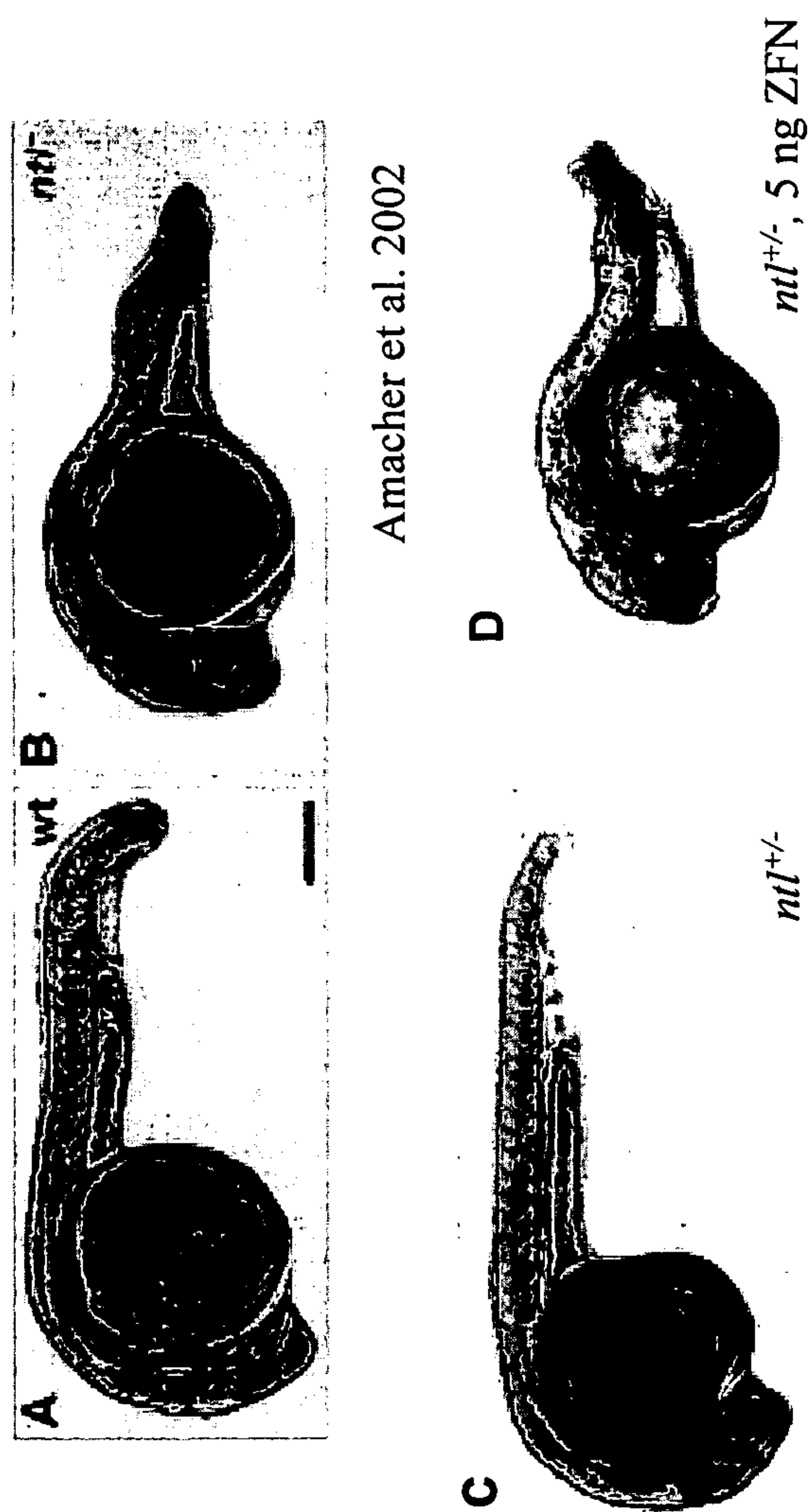


Fig. 12

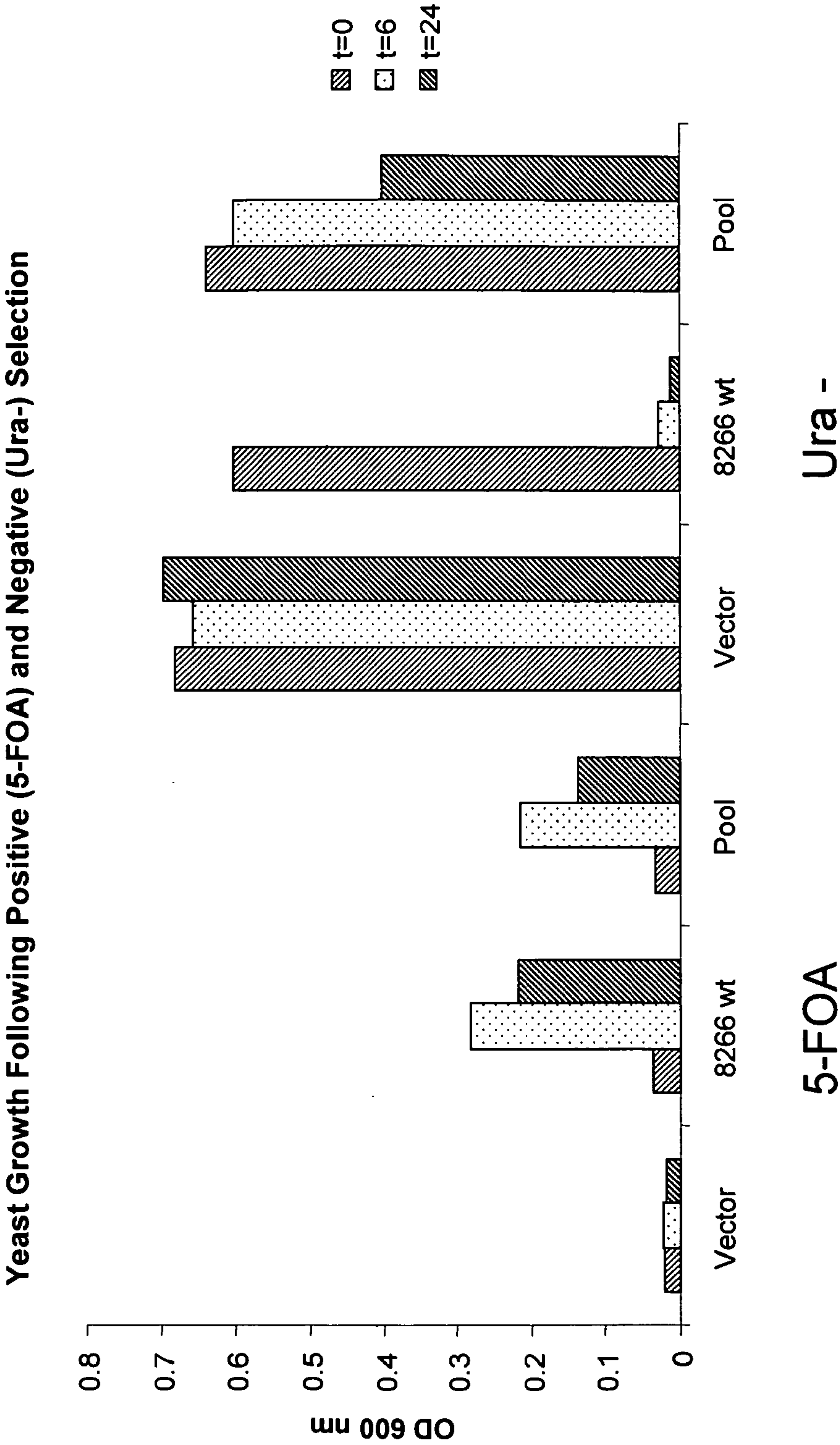


Fig. 13

