

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

(19) World Intellectual Property

Organization

International Bureau

(43) International Publication Date

17 June 2021 (17.06.2021)



(10) International Publication Number

WO 2021/116394 A1

(51) International Patent Classification:

C12N 15/09 (2006.01) A01K 67/027 (2006.01)
C12N 15/63 (2006.01)

MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

(21) International Application Number:

PCT/EP2020/085736

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(22) International Filing Date:

11 December 2020 (11.12.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

19383092.4 11 December 2019 (11.12.2019) EP

(71) Applicant: **CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS (CSIC)** [ES/ES]; C/ Serrano, nº 117, 28006 Madrid (ES).

(72) Inventors: **MARTÍNEZ MORALES, Juan Ramón**; CENTRO ANDALUZ DE BIOLOGIA DEL DESARROLLO, Control genético de la morfogénesis epitelial, Ctra. de Utrera Km.1, 41013 Sevilla (ES). **CORBACHO SÁNCHEZ, Jorge**; CENTRO ANDALUZ DE BIOLOGIA DEL DESARROLLO, Control genético de la morfogénesis epitelial, Ctra. de Utrera Km.1, 41013 Sevilla (ES).

(74) Agent: **PONS ARIÑO, Angel**; PONS Patentes Marcas Internacional S.L., Glorieta Rubén Darío 4, 28010 Madrid (ES).

(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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,

(54) Title: VECTORS FOR TISSUE SPECIFIC TRANSCRIPTOMICS

(57) Abstract: The present invention relates to nucleic acid vectors, methods, uses and kits thereof for the generation of transgenic cells or non-human organism which are suitable for transcriptomic analysis. This disclosure provides transposon-based vectors which contain genetic elements which allow a genetic cassette to be incorporated into the genome of a host cell or a non-human organism. The vectors, methods, uses and kits disclosed herein allow the systematic generation of collection of lines of cells or animals suitable for dynamic and tissue-specific transcriptomic interrogation in different developmental stages or in different physiological situations. The resulting collection of lines will constitute a unique transcriptomics resource for a broad community working in developmental genetics, organ physiology or disease models.



WO 2021/116394 A1

Vectors for tissue specific transcriptomics

DESCRIPTION

5 The invention relates to a novel vector useful in the field of transformation, methods, uses and kits thereof. More particularly, the present invention relates to a modified transposon-based vector that allows the introduction of genetic elements into the genome of cells or non-human organism which facilitate transcriptomics analysis.

10 BACKGROUND ART

An essential step in understanding the complexity of biological systems is understanding the gene expression dynamics of individual cells, or groups of cells (like tissues or organs), i.e., which genes are activated or repress in such cells at a given point in time, both at a basal level and in response to stimuli. The transcriptional status of individual cells, achieved by the unique combinatorial activation and repression of specific genes, is an intrinsic property of each cell type and tissue type. Gene expression dynamics and specificity determine the fate, shape, mechanical properties, metabolic status, physiology, pathology, and survival of each cell. Thus, addressing the question of how individual transcriptomes vary through time and among cell types is of paramount importance in most life science disciplines.

Next Generation Sequencing has revolutionised the study of transcriptomics, the study of all RNAs expressed by an organism, by the development of RNA sequencing (RNA-seq) approach, which enables a global analysis of the transcriptome of cells, tissues and organisms. RNA-seq brought several advantages regarding previous methods, namely high reproducibility, unbiased detection of transcripts, single nucleotide resolution and quantitative estimation of transcript levels over a large dynamic range of expression. Despite its advantages, RNA-seq has been seriously limited by the acquisition of RNA samples from heterogenous cell populations, making the study of transcriptome of specific cell populations a challenge. This is particularly the case when the cell population targeted represents a small fraction of the tissue or when the level of cell heterogeneity is high in the organ of study (i.e. the brain). In this situation, the expression profiling of a given cell population may be mask by the transcriptional “noise” associated to its neighbours. The most commonly used strategy to bypass this

problem is to extract RNA from a homogeneous cell suspension obtained by fluorescence-activated cell sorting (FACS). Unfortunately, cell enrichment by FACS entails long and severe dissociation procedures that may distort gene expression, among other things by inducing a cellular response to stress.

5

To overcome such limitations, a technique to selective labelled translating RNAs in specific cells or groups of cells was developed: Translating Ribosome Affinity Purification (TRAP). TRAP is based on the transformation of a cell or animal with a fusion gene that jointly expresses a ribosome target gene labelled with a marker gene
10 in a specific cell/tissue/organ. The translated fusion protein will get incorporated into translating ribosomes and therefore such ribosomes will also be labelled. This allows to easily isolate the specifically labelled ribosomes from whole organism's extracts by immunoprecipitation, also obtaining the mRNA attached to them, which facilitates the generation of specific transcriptomes of cells/tissues/organs at different developmental
15 stages. The main drawback of the TRAP technique is the difficulty, and at times the impossibility, of having the fusion gene being expressed specifically in certain cells/tissues/organs since nucleic acid elements that specifically promote expression in such targets (like promoters and upstream activating sequences) may be unknown or lack an adequate level of expression for TRAP. In addition, each different target of
20 study requires the generation of transgenic cell/animal, a process which can be cumbersome, difficult and expensive, limiting the application of TRAP (Tryon, R.C. et al. Genesis. 2013. 51(3):187-192).

DESCRIPTION OF THE INVENTION

25

The present invention discloses novel genetic constructions, vectors, methods for its production and uses thereof, which greatly facilitate the creation of libraries of transgenic non-human organism lines which are a valuable resource for the study of transcriptomics, developmental genetics, organ physiology or disease models.

30

The vectors of the present invention solve the problems associated with TRAP by allowing a faster, cheaper, efficient, flexible and consistent generation of transgenic non-human organisms, wherein said organisms have specific cells, tissues or organs labelled, ideal for performing TRAP. The genetic constructions and vectors comprising
35 these genetic constructions, comprises a nucleotide sequence containing several

elements that can randomly integrate into the genome of a target host cell, which will only be active if the integration happens in the proximity of an expression promoting sequence (e.g. an enhancer). Such cells can be used to generate transgenic organisms by known means in the art. The libraries of transgenic host cell and non-human transgenic organism lines created will allow not only for TRAP studies by RNA-seq but also for developmental genetics, physiology, toxicology and biomedical studies. Thus, the genetic constructions and vectors of the present invention are useful to generate transgenic lines suitable for dynamic and tissue-specific transcriptomic interrogation. The resulting collection of lines may constitute a unique resource for a broad community working in developmental genetics, organ physiology or disease models.

As an example of this advantages, Fig. 1 shows a line of zebrafish obtained with the vector of the invention, particularly with the vector shows in Fig. 2A where the tissue lateral line is labelled specifically. It is important to note that is the first time that this tissue has been labelled in such a way, which will allow for studies on the differentiation program and migratory properties of this organ, a reference model for collective cell migration. This is a key process involved in the morphogenesis of diverse organ systems and it has been also identified in wound healing (Scarpa, E., Mayor R. J. Cell. Biol. 2016; 212:143–155). Moreover, collective cell migration is a capacity observed in cancer invasion and metastasis (Lintz, M. et al. J. Biomech. Eng. 2017;139:021005). Therefore, this line can be extremely useful to gain insight into these processes and study mechanisms as cell-cell cohesion, intercellular communication, response to chemotactics or the role of tissue-derived factors that guide the collective migration.

Another example of the benefices of the genetic construction and the vector disclosed herein are the labelling of the entire nervous system as it is shown in Fig. 3E, which includes several neuromeres, but which are globally labelled allowing the analysis of the whole organ. This line makes it possible to perform analysis to investigate how the whole organ is developed, involving the formation of neuromeres, which can be very useful for neurogenesis studies as induction of neural progenitors, specification and differentiation to neurons.

Other example of the utility and versatility of the genetic construction is the specific labelling of the optic cup Fig. 3S, S'. The anatomy and physiology of the eye is very

conserved in most classes of vertebrates. The zebrafish eye has become a great model for studies on eye behaviour and morphogenesis as well as investigation on human eye diseases basis (Richardson, R. et al. Eye. 2017;31:68–86).

- 5 Thus, the invention also relates to a non-human cell and non-human organism which comprises the above-mentioned genetic construction or vector. Moreover, the invention relates to the use of said non-human cell or non-human organism for identification and/or isolation of genes and for the creation of transgenic organisms to study gene function at various developmental stages, including the adult, as well as for the creation
10 of animal models of animal and/or human diseases useful for drug target validation.

In a first aspect, the present invention relates to a genetic cassette, hereinafter genetic cassette of the invention, comprising

- a. a pair of transposase binding sites delimiting said genetic cassette, and
15 b. a fusion gene comprising
i. a reporter gene, and
ii. a ribosomal gene

wherein the fusion gene is in operational combination with a minimal promoter.

- 20 The term "genetic cassette" as used herein refers to any polynucleotide sequence having a discrete and identifiable structure, a fragment or segment of DNA containing a particular grouping of genetic elements. The genetic cassette may be inserted into the target genome. In addition, a genetic cassette can include one or more nucleic acid sequences. An example of a genetic cassette is one that contains at least one,
25 preferably two transposase binding sites, at least one promoter element and one or more nucleic acid for expression. Additionally, the genetic cassette can also comprise a polyadenylation sequence.

- In a specific embodiment, the presence of two transposase binding sites in the
30 expression cassette of the invention means that the genetic cassette has the capacity of insertion into the genome of a host cell upon introduction into said host cell, and therefore the genetic cassette can be replicated along with the genome of said host cell.

In another specific embodiment, the nature of the polyadenylation signal is not believed to be crucial to the successful practice of the methods of the present disclosure, and any such sequence may be employed, including but not limited to the SV40 early polyadenylation signal, the SV40 late polyadenylation signal, the HSV thymidine kinase polyadenylation signal, the protamine gene polyadenylation signal, the adenovirus 5 E1b polyadenylation signal, the bovine growth hormone polyadenylation signal, the human variant growth hormone polyadenylation signal and the rabbit beta-globin polyadenylation signal.

10 In a preferred embodiment, the polyadenylation signal is SV40polyA or functional variants and fragments thereof. In more preferred embodiments the polyadenylation signal is from the polyomavirus simian virus 40 (SV40). In a specific embodiment, the SV40polyA comprise a nucleotide sequence having at least 90% sequence identity with the SEQ ID NO: 1, preferably a 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% identity
15 with the SEQ ID NO: 1. In a more preferred embodiment, the SV40polyA comprise the SEQ ID NO: 1. In a more preferred embodiment the SV40polyA consisting of the SEQ ID NO: 1.

The term "identity", as used in this specification, refers to the proportion of identical amino acids or nucleotides between two compared peptides/proteins or nucleotide sequences, respectively. The methods for comparing sequences are known in the state of the art, and include, but not limited to, the programs BLASTP or BLASTN, ClustalW and FASTA. We can consider that peptides, proteins or nucleotide sequences with percent identities of at least 90% will maintain the same properties as the sequence to
20 which they refer.

In a more preferred embodiment, the genetic cassette of the invention can be removed and inserted into a vector or plasmid as a single unit.

30 The term "nucleic acid" used herein refers to nucleic acid molecules including DNA (gDNA, cDNA), oligonucleotides (double or single stranded), RNAs (sense RNAs, antisense RNAs, mRNAs, tRNAs, rRNAs, small interfering RNAs (siRNAs), double-stranded RNAs (dsRNA), short hairpin RNAs (shRNAs), piwi-interacting RNAs (PiRNA), micro RNAs (miRNAs), small nucleolar RNAs -(SnoRNAs), small nuclear
35 RNAs (SnRNAs)), ribozymes, aptamers, DNazymes, ribonuclease-type complexes and

other such molecules as herein described. For the avoidance of doubt, the term "nucleic acid" includes non-naturally occurring modified forms, as well as naturally occurring forms.

5 A "transposase binding sequence" or "transposase binding site" as used herein refers to the nucleotide sequences that delimit a genetic cassette where a transposase specifically binds when mediating transposition. The transposase binding site may comprise more than one site for binding transposase subunits. Several transposase binding sites are known in the art and a skilful person in the subject matter will be able
10 to identify an ideal transposase binding site for use in the present invention.

A "transposition reaction" as used herein refers to a rearrangement reaction where a genetic cassette is mobilized from a donor nucleic acid into a target genome. Primary components in a transposition reaction are a genetic cassette and a transposase or an
15 integrase enzyme.

The term "transposase" as used herein refers to an enzyme that is a component of a functional nucleic acid-protein complex having the ability to recognize and bind to one end of a genetic cassette or end sequences of a genetic cassette in a rearrangement
20 reaction to promote the mobilization of the genetic cassette. The term "transposase" also refers to integrases from retrotransposons or of retroviral origin.

In a preferred embodiment, the genetic cassette of the present invention comprises a pair of transposase binding sites recognised by a transposase selected from a list
25 consisting of: AC7, Tn5SEQ1, Tn916, Tn951, Tn1721, Tn2410, Tn1681, Tn1, Tn2, Tn3, Tn4, Tn5, Tn6, Tn9, Tn10, Tn30, Tn101, Tn903, Tn501, Tn1000 ($\gamma\delta$), Tn1681, Tn2901, AC/Ds transposons, Mp transposons, Spm transposons, En transposons, Dotted transposons, Mu transposons, dSpm transposons, I transposons, Mariner, Tc3, Mos1, Sleeping Beauty and Tol2. More preferably, the transposase binding sites are
30 recognized by a transposase selected from a list consisting of: Mariner, Tc3, Sleeping Beauty and Tol2, more preferably, Sleeping Beauty and Tol2, and more preferably Tol2.

In a specific embodiment, the Sleeping Beauty transposase comprise an amino acid
35 sequence having at least 90%, sequence identity with the SEQ ID NO: 2, preferably a

91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% identity with the SEQ ID NO: 2. In a more preferred embodiment, the Sleeping Beauty transposase comprises the SEQ ID NO: 2. In a more preferred embodiment the Sleeping Beauty transposase consisting of the SEQ ID NO: 2.

5

In a specific embodiment, the Tol2 transposase comprise an amino acid sequence having at least 90% sequence identity with the SEQ ID NO: 3, preferably a 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% identity with the SEQ ID NO: 3. In a more preferred embodiment, the Tol2 transposase comprises the SEQ ID NO: 3. In a more preferred

10 embodiment the Tol2 transposase consisting of the SEQ ID NO: 3.

In a preferred embodiment, the pair of transposase binding sites are recognized by the Tol2 transposase as described above. Thus, in a preferred embodiment the pair of transposase binding sites comprises the nucleotide sequence SEQ ID NO: 4 as first

15 Tol2 transposase recognition sequence and SEQ ID NO: 5 as second Tol2 transposase recognition sequence.

The term "fusion gene" as used herein refers to the coding regions of two or more genes are connected end to end and placed under the control of the same set of

20 regulatory sequences (including promoters, enhancers, ribosome binding sequences, terminators, etc.), i.e. a chimeric gene or a chimeric DNA. The expression product of the fusion gene is a fusion protein. In certain embodiments the fusion gene in addition comprises a polyadenylation signal, as described above. As it is mentioned previously, the fusion gene of the genetic cassette of the invention comprises a reporter gene and

25 a ribosomal gene.

The term "reporter gene" as used herein refers to a nucleic acid encoding an identifying factor that is able to be identified based upon the reporter gene's effect, wherein the effect is used to track the inheritance of a nucleic acid of interest, to identify a cell,

30 tissue, organ or organism that has inherited the nucleic acid of interest, and/or to measure gene expression induction or transcription. Selectable marker genes may also be considered reporter genes. Examples of reporter genes known and used in the art can be selected from known database such as FPbase (<https://www.fpbase.org>) or similar.

In a particular embodiment, examples of reporter genes useful in the genetic cassette of the present invention include: luciferase (Luc), green fluorescent protein (GFP), enhanced green fluorescent protein (EGFP) red fluorescent protein (RFP), Cyan fluorescent protein (CFP), Yellow fluorescent protein (YFP), chloramphenicol acetyltransferase (CAT), β -galactosidase (LacZ), β -glucuronidase (Gus), and the like. In a more preferred embodiment, the marker gene is GFP, more preferably EGFP. In a more preferably embodiment, the EGFP gene comprises a nucleotide sequence having at least 90% sequence identity with the SEQ ID NO: 6, preferably a 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% identity with the SEQ ID NO: 6. In a more preferred embodiment, the EGFP gene comprises the SEQ ID NO: 6 and more preferably the EGFP gene consisting of the SEQ ID NO: 6. Additionally, the EGFP gene according to the present invention encodes for the EGFP protein comprising an amino acid sequence having at least 90% sequence identity with the SEQ ID NO: 7, preferably a 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% identity with the SEQ ID NO: 7. In a more preferred embodiment, the EGFP comprises the SEQ ID NO: 7 and more preferably the EGFP consisting of the SEQ ID NO: 7.

The term "ribosomal gene" as used herein refers to a gene that expresses a protein that binds to, or associates to, or is part of the ribosome. In some embodiments, the gene is a ribosomal protein gene. In some embodiments, the gene is a large ribosomal subunit protein gene. In some embodiments, the gene is a small ribosomal subunit protein gene. In specific embodiments, the gene is the ribosomal protein L10a gene (rpL10a). In a specific embodiment the ribosomal protein L10a gene is endogenous to the target cell. In a specific embodiment the ribosomal protein L10a is exogenous to the target cell. In a particular embodiment the ribosomal protein L10a gene belongs to a species selected from: mammals, avian, amphibian and fish. In a more preferred embodiment, the ribosomal protein L10a gene belongs to rodents or fish, more particularly belong to mouse or to zebrafish.

In a more particular embodiment, the ribosomal protein L10a gene from mouse comprises a nucleotide sequence having at least 90% sequence identity with the SEQ ID NO: 8, preferably a 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% identity with the SEQ ID NO: 8. In a more preferred embodiment the ribosomal protein L10a from mouse comprises the SEQ ID NO: 8 and more preferably consists of the SEQ ID NO: 8. In a more particular embodiment, the ribosomal protein L10a gene according to the present

invention, encodes for the rpL10a protein comprising an amino acid sequence having at least 90% sequence identity with the SEQ ID NO: 9, preferably a 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% identity with the SEQ ID NO: 9. In a more preferably embodiment the ribosomal protein L10a from mouse comprises the SEQ ID NO: 9 and
5 more preferably consists of the SEQ ID NO: 9.

In a more particular embodiment, the ribosomal protein L10a gene from zebrafish comprises a nucleotide sequence having at least 90% sequence identity with the SEQ ID NO: 10, preferably a 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% identity with the SEQ
10 ID NO: 10. In a more preferably embodiment the ribosomal protein L10a gene from zebrafish comprises the SEQ ID NO: 10 and more preferably consists of the SEQ ID NO: 10. In a more preferred embodiment, the ribosomal protein L10a gene according to the present invention encodes the rpL10a protein which comprises an amino acid sequence having at least 90% sequence identity with the SEQ ID NO: 11, preferably a
15 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% identity with the SEQ ID NO: 11. In a more preferably embodiment the ribosomal protein L10a from zebrafish comprises the SEQ ID NO: 11 and more preferably consists of the SEQ ID NO: 11.

In a more preferred embodiment, the fusion gene of the genetic cassette comprising as
20 reporter gene, the nucleotide sequence comprising the SEQ ID NO: 6 which encodes for the EGFP protein (SEQ ID NO:7), as the ribosomal gene, the ribosomal protein L10a gene from zebrafish which comprises the nucleotide sequence comprising the SEQ ID NO: 10 which encodes for the rpL10a protein (SEQ ID NO: 11). In a more preferred embodiment, the fusion gene comprises the SEQ ID NO: 12.

25
The term "operational combination" as used herein refers to that separate nucleic acid sequences are functionally associated such that an event at one can precipitate a response from the other. Two or more operably linked nucleic acid sequences can, in combination, comprise an independent genetic element, such as an expression
30 cassette.

The term "promoter" as used herein refers to a polynucleotide required for transcription at significant levels of a second polynucleotide to which it is operably combined.

The term "minimal promoter" as used herein refers to promoter that is sensitive, i.e., capable of responding efficiently to different enhancers, and specific, i.e., not able to drive any pattern of expression on its own. Specifically, a minimal promoter refers to a short DNA segment which is typically inactive by itself, but can mediate strong transcription when combined with other transcription regulatory elements. A minimal promoter comprises a transcription start site and functional sequences for binding the transcription start complex (TATA-box) inside a cell, a host cell, or a host organism. Several minimal promoters are known in the art and a skilful person in the subject matter will be able to identify an ideal minimal promoter for use in the specific application.

In a preferred embodiment the minimal promoter is selected from a list consisting of: ef1a, gata1 promoter (gata1p), gata2 promoter (gata2p), keratin8 (krt8), keratin4 (krt4), hsp70, c-fos, E1b, thymidine kinase (TK), carp β -actin and Hspa1a minimal promoters. In a more preferred embodiment, the minimal promoter is the gata2 promoter, more particularly the gata2 promoter belong to *Dario rerio* (zebrafish), and more particularly, comprise a nucleotide sequence having at least 90% sequence identity to SEQ ID NO: 13, preferably a 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% identity with the SEQ ID NO: 4. In a preferred embodiment, the gata2 promoter belong to *Dario rerio* comprising the SEQ ID NO: 13 and more preferably, consisting of the SEQ ID NO: 13.

In a more preferred embodiment, the genetic cassette of the invention comprises the pair of transposase binding sites correspond to the transposase Tol2 and the fusion gene comprises the EGFP as reporter gene and the ribosomal protein L10a gene from zebrafish as ribosomal gene and the Gata2p from zebrafish as a minimal promoter. Thus, in a more preferred embodiment the genetic cassette of the invention comprises the nucleotide sequence SEQ ID NO: 14.

In a more preferred embodiment, the genetic cassette of the invention further comprising a pair of specific recombination sites, preferably a pair of Cre recombinase recognition sites wherein the first Cre recombinase recognition site is located upstream of the minimal promoter and the second Cre recombinase recognition site is located downstream of the fusion gene.

As used herein, "site specific recombination sites" means a pair of DNA segments that is recognized by a site-specific recombinase in a process that allows the excision of the DNA between the pair of DNA segments. For instance, Cre recombinase recognizes either loxP recombination sites or lox511 recombination sites which are hetero-specific, 5 which means that loxP and lox511 do not recombine together. The Cre/lox system is disclosed by Odell et al., Plant Physiol, 106(2): 447-58 (1994).

The term "Cre recombinase" as used herein refers to the site-specific DNA recombinase derived from the Pi bacteriophage. Cre recombinase is a 38-kDa product 10 of the cre (cyclization recombination) gene of bacteriophage Pi and is a site-specific DNA recombinase of the nt family (Sternberg, N. et al. J. Mol. Biol. 1986. 187:197-212). Cre recombinase recognizes a 34-bp sequence on the Pi genome called LoxP (locus of X-over Pi, LoxP site) and efficiently catalyses reciprocal conservative DNA recombination between pairs of LoxP sites. Cre recombinase-mediated recombination 15 between two directly repeated LoxP sites results in excision of DNA between.

The term "lox site" as used herein refers to a nucleotide sequence at which the product of the Cre gene of bacteriophage P1, the Cre recombinase, can catalyse a site-specific recombination event. A variety of lox sites are known in the art, including the naturally 20 occurring loxP, loxB, loxL and loxR, as well as a number of mutant, or variant, lox sites, such as loxP511, loxP514, lox Δ 86, lox Δ 117, loxC2, loxP2, loxP3 and lox P23.

In a specific embodiment, the pair of Cre recombinase recognition sites, particularly the pair of LoxP sites, comprises a nucleotide sequence having at least 90% identity to the 25 SEQ ID NO: 15, preferably a 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% identity with the SEQ ID NO: 15. In a further preferred embodiment, the pair of LoxP sites comprises the SEQ ID NO: 15, more preferably consisting of SEQ ID NO: 15.

An alternative recombinase is a FLP recombinase. As used herein, the term "FLP 30 recombinase" refers to a site-specific DNA recombinase derived from yeast, the 2pi plasmid of *Saccharomyces cerevisiae*, that recognizes a 34 base pair DNA sequence, termed the "FRT site" (FLP recombinase target). It is also recognized by one skilled in the art that modified forms or mutants of FLP recombinase can recognize the FRT site and its variants. Therefore, the term "FLP recombinase" and "FRT site" includes all 35 variants and mutants carrying out their functions as described above.

As used herein, the terms "upstream" and "downstream" are used to differentiate relative positions in DNA or RNA sequences. The upstream position is a position towards the 5' end of another nucleic acid segment (e.g. a promoter, gene, restriction site, etc.) in a single-stranded DNA or RNA molecule. The downstream position is a position towards the 3' of another segment of nucleic acid in a single stranded DNA or RNA molecule.

In another preferred embodiment, the genetic cassette of the invention further comprising a nucleotide sequence for an integrase specific site.

The term "integrase specific site" or "ISS" as used herein means *attP* or *attB*. When a first ISS is *attP*, a corresponding second ISS is *attB* (or the first ISS is *attB* and the second ISS is *attP*) such that the first ISS and the second ISS can undergo site specific integration mediated by or in presence of an integrase. In a preferred embodiment the ISS is located outside the region delimited by the Cre recombinase recognition sites. Preferably, site specific integration is mediated by the action of integrases which act unidirectional. Said integrases therefore increase the likelihood of persistent integration as they cannot also mediate excision of the integrated sequence in contrast to bidirectional recombinases. Accordingly, in a preferred embodiment the ISS is a sequence recognizable by an integrase.

In another preferred embodiment the integrase specific site is *attP* and comprises the nucleotide sequence SEQ ID NO: 16.

In a more preferred embodiment, the genetic cassette of the invention comprises the pair of transposase binding sites correspond to the transposase Tol2, the fusion gene comprises the EGFP as reporter gene and the ribosomal protein L10a gene from zebrafish as ribosomal gene, the Gata2p from zebrafish as the minimal promoter, the LoxP as the Cre recombinase recognition site and an integrase specific site. Thus, in a more preferred embodiment, the genetic cassette of the invention comprises the nucleotide sequence SEQ ID NO: 17.

A further aspect of the present invention relates to a vector, hereinafter the vector of the invention, comprising the genetic cassette of the invention.

The term "vector" as used herein refers to a nucleic acid molecule capable of transporting one or more other nucleic acid sequence(s), like a genetic cassette, to which it has been linked or which was introduced into said vector. One type of vector is a "plasmid", which refers to a circular double stranded DNA into which additional DNA segments may be cloned. Another type of vector is a viral vector, wherein additional DNA segments may be cloned into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) or parts thereof, are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" or simply "expression vectors".

In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably, as the plasmid is the most commonly used form of vector. However, the disclosure is intended to include other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions. The vector may also contain additional sequences, such as a "polylinker" or "multiple cloning site" for subcloning of additional nucleic acid sequences. A "vector backbone" or "plasmid backbone" refers to a piece of DNA containing at least a plasmid origin of replication and a selectable marker which allows for selection of hosts cells containing the plasmid or vector. When a particular plasmid or vector is modified to contain non-plasmid elements (e.g. insertion of Ad sequences and/or a eukaryotic gene of interest linked to a ribosomal promoter), the plasmid sequences are referred to as the plasmid backbone.

The term "selectable marker" has used herein means an identifying factor, usually an antibiotic or chemical resistance gene, that is able to be selected for based upon the marker gene's effect, i.e., resistance to an antibiotic, resistance to a herbicide, colorimetric markers, enzymes, fluorescent markers, and the like, wherein the effect is used to track the inheritance of a nucleic acid of interest and/or to identify a cell or organism that has inherited the nucleic acid of interest. Examples of selectable marker genes known and used in the art include: genes providing resistance to ampicillin,

streptomycin, gentamycin, kanamycin, hygromycin, bialaphos herbicide, sulfonamide, and the like; and genes that are used as phenotypic markers, i.e., anthocyanin regulatory genes, isopentanyl transferase gene, and the like.

- 5 Vectors may be introduced into the desired host cells by methods known in the art, e.g., transfection, electroporation, microinjection, transduction, cell fusion, DEAE dextran, calcium phosphate precipitation, lipofection (lysosome fusion), use of a gene gun, or a DNA vector transporter (see, e.g., Wu et al., J. Biol. Chem. 1992, 267:963-967; Wu and Wu, J. Biol. Chem. 1988, 263:14621-14624; and Hartmut et al., Canadian
10 Patent Application No. 2,012,311, filed March 15, 1990).

The term "transfection" means the uptake of exogenous or heterologous RNA or DNA by a host cell. A cell has been "transfected" by exogenous or heterologous RNA or DNA when such RNA or DNA has been introduced inside the host cell. A cell has been
15 "transformed" by exogenous or heterologous RNA or DNA when the transfected RNA or DNA effects a phenotypic change. The transforming RNA or DNA can be integrated (covalently linked) into chromosomal DNA making up the genome of the cell.

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a
20 host cell or organism, resulting in genetically stable inheritance. Host cells or organisms containing the transformed nucleic acid fragments are referred to as "transgenic" or "recombinant" or "transformed" organisms.

In addition, the vector of the invention comprising the cassette of the invention may
25 include one or more origins for replication in the cellular hosts in which their amplification or their expression is sought, reporter genes or selectable markers.

In a particular embodiment, the vectors of the present invention may be episomal or
30 integrating.

In a preferred embodiment, the vector of the invention comprises the genetic cassette which comprises the SEQ ID NO: 14. In another preferred embodiment, the vector of the invention comprises the genetic cassette which comprises the SEQ ID NO: 17. In a more preferred embodiment, the vector of the invention is selected from the SEQ ID
35 NO: 18 or the SEQ ID NO: 19.

In addition to the aforementioned characteristics, the genetic cassette and the vectors of the present invention also allow for further studies, non-TRAP related, to be performed, since further embodiments of the present invention allow for the partial
5 removal of the genetic cassette as well as targeted integration into the genome at the locus of the genetic cassette.

More particularly, the genetic cassette and the vectors of the present invention are useful:

- 10
- (i) for generating host cells and/or non-human organisms,
 - (ii) for tissue-specific transcriptomics analysis,
 - (iii) to study gene function at various developmental stages,
 - (iv) as a model of animal and/or human diseases, or
 - (v) as a drug validation model in drug development.

15

A further aspect of the invention relates to a host cell, hereinafter host cell of the invention, comprising the genetic cassette of the invention and/or the vector of the invention.

20 The term "host cell" as used herein refers to a cell into which exogenous DNA has been introduced, either *in vitro*, *ex-vivo*, or *in vivo*. It should be understood that such terms are intended to refer not only to the particular subject cell, but, also to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not,
25 in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein. A host cell can be used, for example, for expression of a nucleic acid of interest, propagation of plasmid vectors and/or delivery of a sequence of interest to cells, wherein the nucleic acid of interest or plasmid vector are present episomally in these cells. The episomal vectors are not integrated into the host genome
30 and are retained in self-replicating form in the host cell. The term episomally replicating is understood here as that the vector is not integrated to the genome of the host cell, but exists in parallel, is also replicated during the cell cycle and in the course of this the vector copies—depending on the number of the copies present before and after cell division—are distributed statistically in the resulting cells.

Examples of such cells include, without limitation, bacterial cells like *Escherichia coli* cells and *Pichia pastoris* cells, yeast cells like *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, insect cells like *Spodoptera frugiperda* Sf21 and Sf19 cells, mammalian cells like human embryonal kidney cells, Chinese hamster ovary cells and Hella cells. In addition, host cells can also be cells obtained by integrating the genetic cassette, or the vector of the invention, into their genome by known methods in the art. Preferably host cells include prokaryotic and eukaryotic cells selected from any of the kingdoms of life. Several cell types are known in the art and a skilful person in the subject matter will be able to identify an ideal cell type for use in the specific application. Preferred cells include, but are not limited to, mammalian cells, avian cells, amphibian cells, plant cells, fish cells, bacterial cells and insect cells. In a more preferred embodiment, the host cell is selected from a list comprising gram-negative bacteria cells, zebrafish cells, *Drosophila* cells, mouse cells and non-human cells.

15 A further aspect of the invention relates to a non-human organism, hereinafter the non-human organism of the invention, comprising the genetic cassette of the invention, the vector of the invention, or the host cell of the invention.

The term "organism" as used herein refers to a body carrying on processes of life, which has various properties, such as, representatively, cellular structure, proliferation (self-reproduction), growth, regulation, metabolism, repair ability, and the like. Typically, organisms possess basic attributes, such as heredity controlled by nucleic acids and proliferation in which metabolism controlled by proteins is involved. Organisms may include natural, wild-type, artificially manipulated, genetically modified, hybridized, or other variants or isolates. Organisms include viruses, prokaryotic organisms, eukaryotic organisms (e.g., unicellular organisms such as yeast, etc.) and multicellular organisms (e.g., plants, animals, etc.). It will be understood that, as the term is used herein, "organism" also refers to and encompasses cells as defined herein, and tissues derived from them, and that the methods of the present disclosure may be applied to any such cell or tissues.

In a specific embodiment, the non-human organism is a prokaryotic organism or a eukaryotic organism. In another specific embodiment, the non-human organism is an invertebrate organism or a vertebrate organism. Preferably, the non-human organism is selected from the group consisting of a bacterium, a fungus, a yeast, a nematode, an

insect, a fish, a plant, a bird, an animal, and a mammal. More preferably, the non-human organism is a yeast, a nematode, an insect, a plant, a fish, an amphibian, a chicken, a hamster, a mouse, a rat, a rabbit, a cat, a dog, a bovine, a goat, a cow, a pig, a horse, a sheep, a simian, a monkey, or a chimpanzee. In a preferred
5 embodiment, the animal is a rodent. In another preferred embodiment, the animal is a fish. In a more preferred embodiment, the animal is a zebrafish.

In another aspect, the invention comprises a kit and/or device comprising the genetic cassette, the vector, the host cell, the culture of host cell, the tissue, the non-human
10 organism of the invention, or any combination thereof.

The kit of the invention comprises at least the genetic cassette of the invention or the vector of the invention having the genetic cassette, said genetic cassette having the necessary elements to allow its integration into the genome of a host cell. Said host cell
15 may in addition also be contained in the kit. The kit further may contain a nucleic acid with transposase activity that allows the integration of the genetic cassette, or from the genetic cassette from the vector into the genome of the host cell. The kit may further include other components that find use in the subject invention, e.g., buffers, delivery vehicles, etc. The various components of the kit may be present in separate containers
20 or certain compatible components may be combined into a single container, as desired. In addition to the above components, the subject kits will further include instructions for practicing the subject invention. These instructions may be present in the subject kits in a variety of forms, one or more of which may be present in the kit. One form in which these instructions may be present is as printed information on a suitable medium or
25 substrate, e.g., a piece or pieces of paper on which the information is printed, in the packaging of the kit, in a package insert, etc. Yet another means would be a computer readable medium, e.g. a CD, an USB, etc., on which the information has been recorded. Yet another means that may be present is a website address which may be used via the internet to access the information at a removed site. Any convenient
30 means may be present in the kits.

Another aspect of the present invention includes a method *in vitro* or *ex vivo* for producing a transgenic host cell and/or a transgenic non-human organism of the invention wherein the method comprises the insertion of the genetic cassette or the
35 vector of the invention into a suitable cell and/or a suitable non-human organism.

Methods to produce a transgenic host cell or a transgenic non-human organism according to the present invention, preferably by using transposon-based nucleic acids such as vectors are well known in the art and several methodologies can be found in the literature (Miskey, C., *et al.* CMLS, Cell. Mol. Life Sci. (2005) 62: 629). The method of present invention may in addition contain other steps like culturing the non-human organism host cells into a fully developed non-human organism – the founder organism, screening for the reporter gene in the founder organism, outcrossing the founder organism with *wild type* non-human organism to generate F1 generation organism, screening the F1 generation for specific tissue labelling of the reporter gene, label of founder organism according to the specific reporter gene label detected in F1 generation, establish stable organism lines from founder organism.

A "transgenic non-human animal" refers to an animal, other than a human, that comprises a gene construct ("transgene") not found in a wild-type animal of the same species or breed. A "transgene" as referred to in this context has the normal meaning in the art of biotechnology and includes a genetic sequence which has been produced or altered by recombinant DNA or RNA technology and which has been introduced into an animal cell. The transgene may include genetic sequences derived from an animal cell, which may be of the same or different species or breed as the cell into which the transgene is introduced. Typically, the transgene has been introduced into the animal by human manipulation such as, for example, by transformation but any method can be used as one of skill in the art recognizes. In the present invention the transgene is the genetic cassette or the vector of the invention.

As it is mentioned previously, techniques for producing transgenic non-human animals are well known in the art. Transformation of a polynucleotide molecule into a cell can be accomplished by any method by which a polynucleotide molecule can be inserted into the cell. Transformation techniques include, but are not limited to, transfection, electroporation, microinjection, lipofection, adsorption, and cell fusion. A recombinant cell may remain unicellular or may grow into a tissue, organ or a multicellular organism. Transformed polynucleotide molecules can remain extrachromosomal or can integrate into one or more sites within a chromosome of the transformed (i.e., recombinant) cell in such a manner that their ability to be expressed is retained. In a preferred embodiment, the transformation technique is microinjection. In a most preferred

method, however, the appropriate genetic cassette or vector of the invention are microinjected into the pronucleus or cytoplasm of embryos, preferably at the single cell stage, and the embryos allowed to develop into mature transgenic animals.

- 5 In another aspect, the present invention also relates to the use *in vitro* or *ex vivo* of the genetic cassette, or the vector of the invention for generating transgenic host cells and/or transgenic non-human organisms.

In another aspect, the present invention also relates to the use *in vitro* or *ex vivo* of the
10 genetic cassette, the vector, the transgenic host cell, the transgenic non-human organism, or the kit according to the present invention,

- (i) for tissue-specific transcriptomics analysis,
- (ii) to study gene function at various developmental stages,
- (iii) as an animal model of human disease, or
- 15 (iv) as a drug validation model in drug development.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skilled in the art to which this invention belongs. Methods and materials similar or equivalent to those described
20 herein can be used in the practice of the present invention. Throughout the description and claims the word "comprise" and its variations are not intended to exclude other technical features, additives, components, or steps. Additional objects, advantages and features of the invention will become apparent to those skilled in the art upon examination of the description or may be learned by practice of the invention. The
25 following examples, drawings and sequence listing are provided by way of illustration and are not intended to be limiting of the present invention.

DESCRIPTION OF THE DRAWINGS

Fig. 1. EGFP-rpl10a expression in the transgenic line TT15 at 35 hours post-fertilization (hpf) (lateral view) is limited to the lateral line system cells (Y = yolk; E =eye).

Fig. 2 Schematic representation of the vectors used in the examples.

5 **Fig. 3.** EGFP-rpl10a fusion gene expression profiles in different zebrafish transgenic lines: TT1 at 48hpf (A, A'), hindbrain (medulla oblongata and spinal cord, respectively); TT2 at 48 hpf (B), 72 hpf (B'), lenses; TT5 at 48hpf (C, C'), first branchial arch and jaw, respectively; TT6 at 48hpf (D), somites; TT7 at 48hpf (E), central nervous system; TT8 at 48hpf (F), 72hpf (F') olfactory bulb; TT9 at 48hpf (G) 72hpf (G') optic tectum; TT10 at
10 48hpf (H) neurons from dorsal root ganglion; TT11 at 48hpf (I), 72hpf (I') branchial arches, jaw and heart, respectively; TT13 at 48hpf (J) dorsal root ganglion; TT15 at 40hpf (K) lateral line system; TT16 at 48hpf (L), 72hpf (L'), mesencephalon; TT18 at 48 hpf (M) whole nervous system; TT21 at 48hpf (N), 72hpf (N') mesencephalon; TT25 at 48 hpf (O) 72 hpf (O') anterior nervous system; TT27 at 5dpf (P, P') telencephalon;
15 TT28 at 48hpf (Q, Q') hindbrain (spinal cord, medulla oblongata, cerebellum) and fins, respectively; TT35 at 48hpf (R, R') telencephalon, eyes, respectively; TT37 at 48hpf (S, S') eyes; TT42 at 48 hpf (T, T') mesencephalon; TT46 at 48 hpf (U, U') telencephalon, mesencephalon, respectively; TT47 at 48hpf (V, V') endoderm, fins, archial branches, respectively; TT50 at 48hpf (W) spinal cord, somites; TT53 at 48hpf (X) 72 hpf (X') fore and midbrain, eyes, respectively; TT54 at 48hpf (Y, Y') fins, hindbrain, branchial arches, respectively; TT60 at 48 hpf (Z), notochord.

The following examples are offered by way of illustration and not by way of limitation. It is understood that the examples and embodiments described herein are for illustrative
25 purposes only, and persons skilled in the art will recognize various reagents or parameters that can be altered without departing from the spirit of the invention or the scope of the appended claims.

EXAMPLES

30

Materials and Methods

Genetic cassette and Vector construction

The vectors of the invention may be produced by standard methods of restriction enzyme cleavage, ligation and molecular cloning as known for the skilled person. Briefly, the vector of the invention named as Tol2_trap:TRAP (SEQ ID NO: 18) was created inserting the minimal promoter gata2p (SEQ ID NO: 13) to the commercial plasmid Tol2-zTRAP from Addgene (plasmid #42236). The gata2 promoter (SEQ ID NO: 13) was extracted from the ZED vector (Bessa, J. *et al.* Developmental dynamics. 2009, 238(9):2409-2417) using the restriction enzymes Sall and BamHI (1h at 37°C). The commercial Tol2-zTRAP vector (Addgene) was also cut with these two restriction enzymes in the same conditions. Afterwards, both elements were ligated using the T4 DNA ligase enzyme, incubating overnight at 16°C. The next day, *Escherichia coli* competent cells (DH5 α) were transformed with the ligation and plated and cultured in LB agar with ampicillin. To select the correctly ligated plasmids, bacterial cultures were prepared in liquid LB from single colonies picked from the plate. The plasmids were isolated from each culture and analysed with a restriction assay using the NheI enzyme. The correct plasmid was purified with a plasmid miniprep kit (GenElute™, Sigma-Aldrich) and its concentration measured by spectrophotometry (Nanodrop One, Thermo Fisher). Further components could be added to the vector/plasmid using similar methods.

20 EGFP Zebrafish Transgenic lines generation

The zebrafish transgenic lines are generated through the insertion of the vector of the invention in the host cells by microinjection of one-cell stage embryos. Briefly, the vector of the invention Tol2_trap:TRAP (SEQ ID NO: 18) carrying the genetic cassette of the invention (SEQ ID NO: 14) can be used. In order to obtain one-cell stage embryos, male and female adult zebrafishes are separated the day before. At the moment of the microinjection they are put together in a breeding tank. After the embryos have been laid, the zebrafishes are returned to their original tank and the embryos are collected using a strainer. The embryos are placed in a petri dish containing saline solution, ready to be microinjected. For this, a needle glass attached to a micromanipulator is used. The micromanipulator is necessary to control the needle and change its position in an easy way. The needle is filled with a solution containing 25ng/ μ L of the purified vector of the invention (SEQ ID NO: 18) and 25ng/ μ L of the Tol2 transposase mRNA (SEQ ID NO: 20). Each embryo in one-cell stage is microinjected with 5nL of the solution. Afterwards the embryos are kept at 28°C and 24

hours post-fertilization (hpf) they are examined using a fluorescence micro stereoscope in search of EGFP-rpl10a expression (seen as green fluorescence). All the positive embryos were selected and raised. For this, they were placed in a petri dish at a maximum number of 100 embryos per plate. 5 days after fertilization the larvae were
5 transferred to regular tanks and fed. During the first few days they were kept in around 3 cm of water, increasing its volume gradually each 2 or 3 days until the larvae are 3 weeks old and they can be in the water system flow. When fishes are adult and sexually mature (around 3 months old) they can breed.

10 From this moment, the grown microinjected fishes are crossed one by one with wild-type ones (outcross). The embryos of each cross are collected and screened for GFP expression at 8, 24, 48 and 72 hpf using a fluorescence micro stereoscope. If less than 150 embryos were obtained from an outcross, an additional one was set up in order to get enough embryos for the screening. The fishes that transmitted GFP expression to
15 their progeny were considered as positives (founders) and labelled with a number that designated the line of which the fish was founder. The embryos that showed tissue-specific GFP expression were raised to establish and maintain each transgenic line.

Imaging

20 To characterize the GFP expression pattern of each transgenic line, pictures of the embryos were taken at different developmental stages. For this, each founder fish was crossed with a wild-type one and the collected embryos were kept at 28°C. At the moment of taking the photographs, the embryos were dechorionated with forceps if
25 needed and mounted in 2% methylcellulose. The photographs were taken at 75% epiboly (8hpf), 24 hpf, 48 hpf and 72 hpf, using a camera Nikon and an Olympus fluorescence micro stereoscope.

Example 1 – Transformation efficiency and germline transmission.

30 One-cell stage zebrafish embryos were injected with the composition comprising the vector of the invention comprising the genetic cassette of the invention shown in Fig. 1A and the Tol2 transposase expressing mRNA, as explained previously in material and methods. The embryos that presented some type of fluorescence were selected to
35 be raised to adulthood using the established and standard husbandry practices

commonly known in the state of the art (Lawrence C. Aquaculture. 2007. Vol.269 (1-4), Pag 1-20).

5 A total of 300 embryos reached adulthood. These embryos were then screened as described above for the marker gene, EGFP fluorescence. From the adult zebrafishes screened, 53 zebrafishes (approx. 17.5%) were positive for the marker gene, showing fluorescence signal of some kind. From these, 33 (approx. 11%) showed tissue-specific EGFP expression, i.e., expression of the marker gene constrained to a specific and delimited region in the zebrafish.

10

The results indicated that the efficiency of transformation was of 17.5% Furthermore, the results also indicate that the integration of the genetic cassette was unbiased and occurred randomly as demonstrated by the diversity of tissues presenting detectable EGFP expression, regardless of the germ layer from which the tissues derive or the
15 tissue position along the embryo axes (see Example 2).

Example 2 - Transgenic lines' characterization.

Embryos from each tissue-specific transgenic line generated were examined with a
20 fluorescent micro stereoscope. Moreover, in order to keep record of the peculiarities of each line, photographs were taken at different stages of development.

The Fig. 3 shows images of the obtained lines having different expression patterns. The obtained transgenic lines has been identified with the letters "TT" followed by a
25 number: TT1 at 48hpf (A, A'), hindbrain (medulla oblongata and spinal cord, respectively); TT2 at 48 hpf (B), 72 hpf (B'), lenses; TT5 at 48hpf (C, C'), first branchial arch and jaw, respectively; TT6 at 48hpf (D), somites; TT7 at 48hpf (E), central nervous system; TT8 at 48hpf (F), 72hpf (F') olfactory bulb; TT9 at 48hpf (G) 72hpf (G') optic tectum; TT10 at 48hpf (H) neurons from dorsal root ganglion; TT11 at 48hpf (I), 72hpf
30 (I') branchial arches, jaw and heart, respectively; TT13 at 48hpf (J) dorsal root ganglion; TT15 at 40hpf (K) lateral line system; TT16 at 48hpf (L), 72hpf (L'), mesencephalon; TT18 at 48 hpf (M) whole nervous system; TT21 at 48hpf (N), 72hpf (N') mesencephalon; TT25 at 48 hpf (O) 72 hpf (O') anterior nervous system; TT27 at 5dpf (P, P') telencephalon; TT28 at 48hpf (Q, Q') hindbrain (spinal cord, medulla oblongata, cerebellum) and fins, respectively; TT35 at 48hpf (R, R') telencephalon,
35

eyes, respectively; TT37 at 48hpf (S, S') eyes; TT42 at 48 hpf (T, T') mesencephalon; TT46 at 48 hpf (U, U') telencephalon, mesencephalon, respectively; TT47 at 48hpf (V, V') endoderm, fins, archial branches, respectively; TT50 at 48hpf (W) spinal cord, somites; TT53 at 48hpf (X) 72 hpf (X') fore and midbrain, eyes, respectively; TT54 at 5 48hpf (Y, Y') fins, hindbrain, branchial arches, respectively; TT60 at 48 hpf (Z), notochord.

Conclusions

10 To date, we have identified 53 founders (zebrafish transgenic lines) that showed EGFP-rpl10a expression. Out of them, 33 showed specific expression patterns (24 lines showed in Fig. 3). These transgenic lines show very specific and unique expression patterns, some of them not previously described (TT5, TT15 or TT21). Therefore, the results show in the present invention indicate that the cassettes and the
15 vectors of the invention in combination with the associated technology will constitute an improved tool for the dynamic analysis of tissue-specific transcriptomes. These unique transgenic lines will constitute powerful experimental tools for groups interested in the development and the physiology of the concerned tissues. Given the very general applicability of RNA-seq studies in biomedicine, and the sustained growth of model
20 systems for the study of animal and/or human diseases, it is believed that the present invention has an important translational applicability. In the specific case of zebrafish, the systematic generation of transgenic lines should be relevant not only in developmental genetics, but also in physiology, toxicology and biomedical studies. Thus, the long-term economic and medical impact of developing these methods should
25 not be underestimated.

CLAIMS

1. A genetic cassette comprising
- 5 a) a pair of transposase binding sites delimiting said genetic cassette, wherein the pair of transposase binding sites is recognized by Tol2 transposase and comprise the SEQ ID NO: 4 and the SEQ ID NO: 5, and
- b) a fusion gene comprising
- 10 i) a reporter gene, and
- ii) a ribosomal gene
- wherein the fusion gene is in operational combination with a minimal promoter.
2. The genetic cassette according to claim 1 wherein the reporter gene is selected from the list consisting of: luciferase (Luc), green fluorescent protein (GFP),
15 enhanced green fluorescent protein (EGFP) red fluorescent protein (RFP), Cyan fluorescent protein (CFP), Yellow fluorescent protein (YFP), chloramphenicol acetyltransferase (CAT), β -galactosidase (LacZ) and β -glucuronidase (Gus), preferably the reporter gene is EGFP wherein the EGFP comprises the SEQ ID NO: 6.
- 20 3. The genetic cassette according to claim 1 or 2 wherein the ribosomal gene is a large ribosomal subunit protein gene or a small ribosomal subunit protein gene, preferably, the ribosomal gene is the ribosomal protein L10a gene comprising the SEQ ID NO: 8 or SEQ ID NO: 10.
- 25 4. The genetic cassette according to any of claims 1 to 3 wherein the minimal promoter is selected from the list consisting of: ef1a, gata1p, gata2p, keratin8, keratin4, hsp70, c-fos, E1b, thymidine kinase, carp β -actin, Hspa1a, preferably, the minimal promoter is the gata2p comprising the SEQ ID NO: 13.
- 30 5. The genetic cassette according to any of claims 1 to 4 further comprising a pair of Cre recombinase recognition sites, wherein the first Cre recombinase recognition site is located upstream of the minimal promoter and the second Cre recombinase recognition site is located downstream of the fusion gene,
35 preferably is a LoxP site comprising the SEQ ID NO: 15.

6. The genetic cassette according to any of claims 1 to 5 further comprising an integrase specific site, preferably an attP site comprises the SEQ ID NO: 16.
- 5 7. A genetic cassette according to any of claims 1 to 6 characterized in that comprises the SEQ ID NO: 14 or SEQ ID NO: 17.
8. A vector comprising the genetic cassette according to any of claims 1 to 7.
- 10 9. A host cell, a culture of host cell, a tissue or a non-human organism comprising the genetic cassette according to any of claims 1 to 7 or the vector according to claim 8.
- 15 10. A kit comprising the genetic cassette according to any of claims 1 to 7, the vector according to claim 8, the host cell, the culture of host cell, the tissue or the non-human organism according to claim 9, or any combination thereof.
- 20 11. An *in vitro* or *ex vivo* method for producing the host cell, the culture of host cell, the tissue or the non-human organism of claim 9, which comprises the insertion of the genetic cassette according to any of claims 1 to 7, or the vector according to claim 8, into a suitable cell.
- 25 12. Use *in vitro* or *ex vivo* of the genetic cassette according to any of claims 1 to 7 or the vector according to claim 8, for generating host cells, culture of host cell, tissue and/or non-human organisms.
- 30 13. Use *in vitro* or *ex vivo* of the genetic cassette according to any of claims 1 to 7, the vector according to claim 8, the host cell, the culture of host cell, the tissue and/or the non-human organism according to claim 9, or the kit according to claim 10
- (i) for tissue-specific transcriptomics analysis,
 - (ii) to study gene function at various developmental stages,
 - (iii) as an animal model of animal and/or human disease, or

(iv) as a drug validation model in drug development.

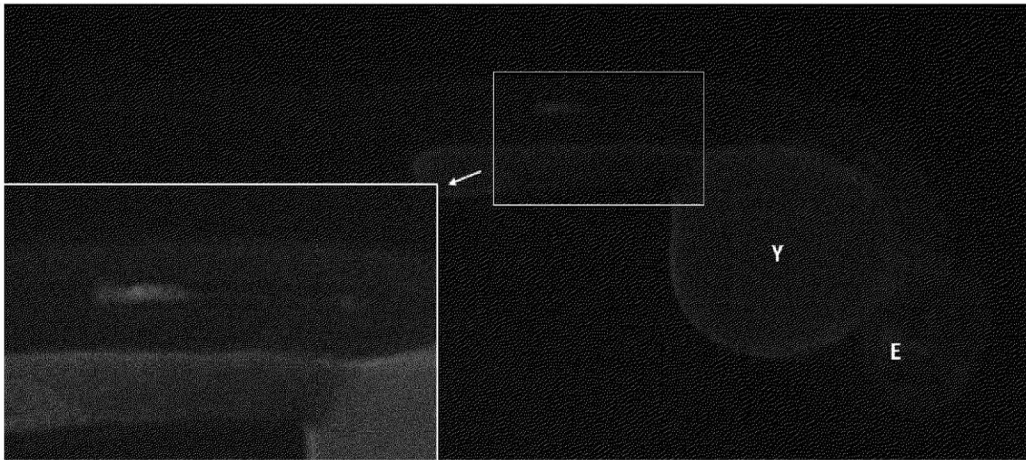


Fig. 1

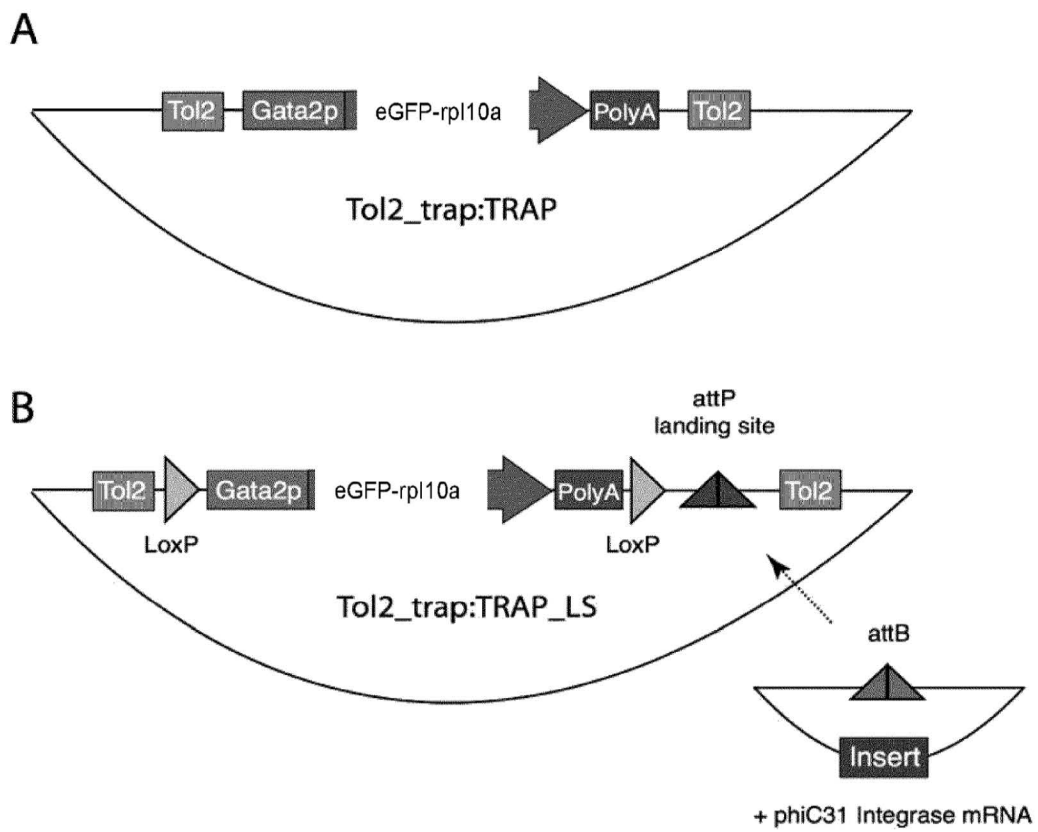


Fig 2.

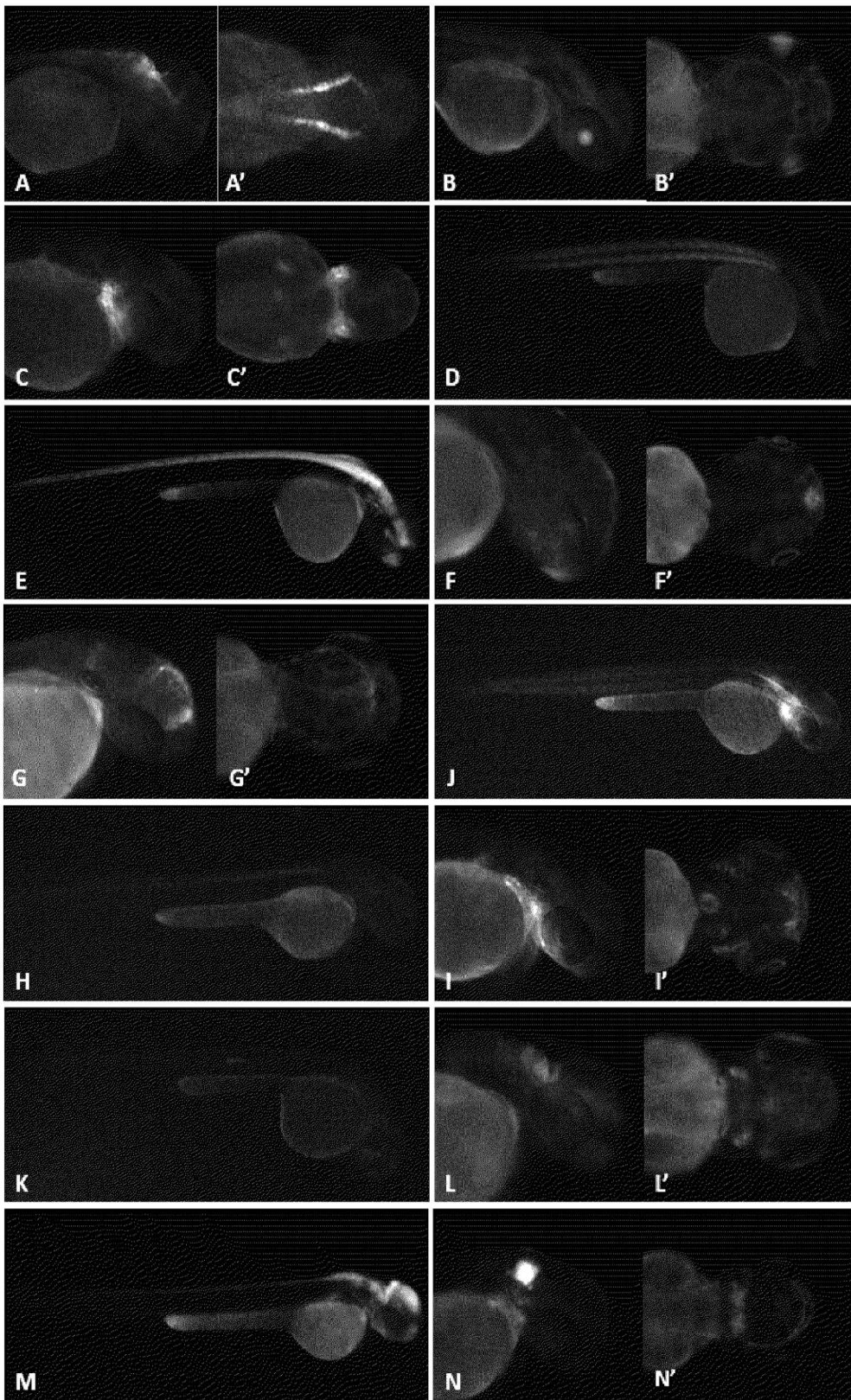


Fig. 3

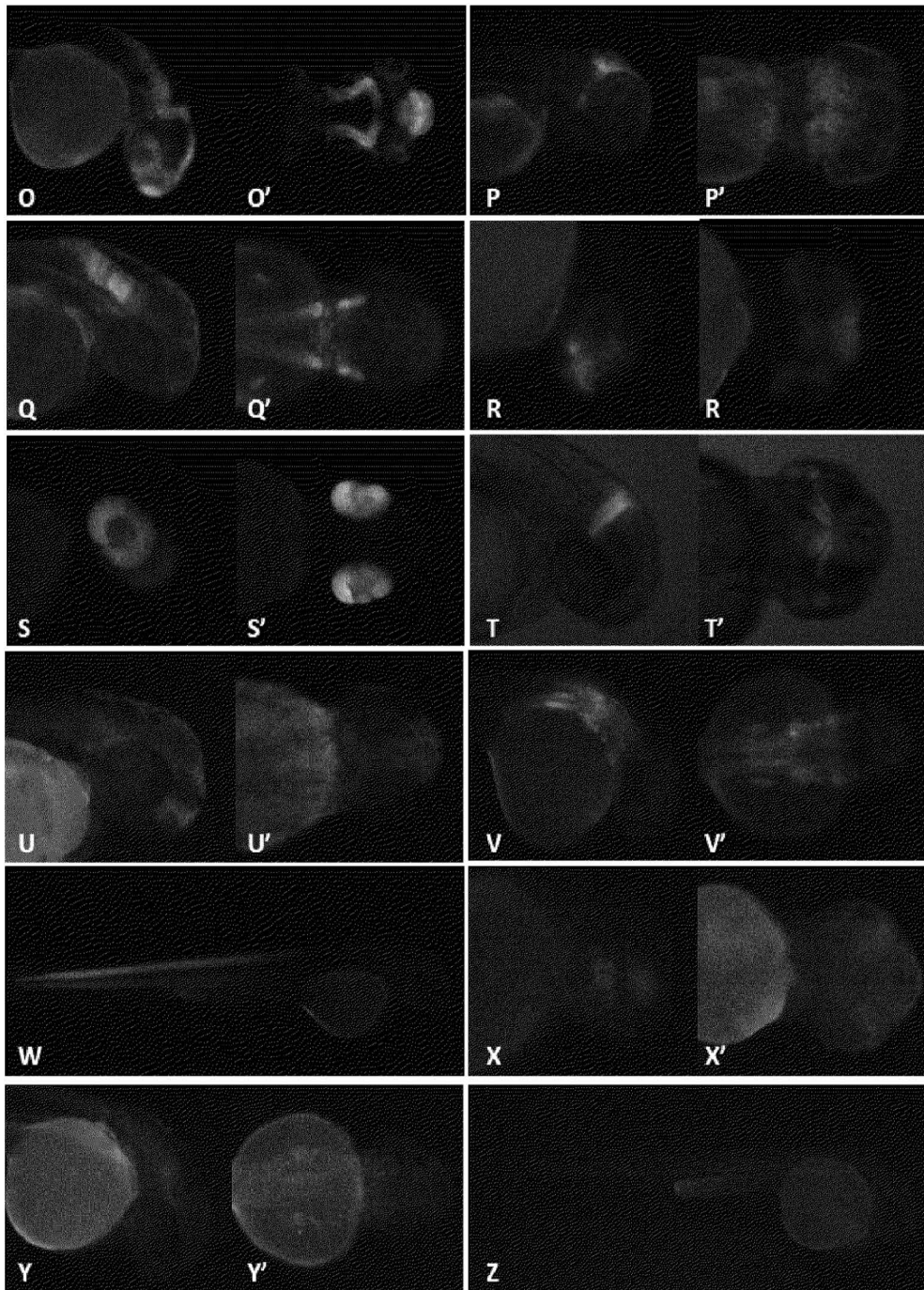


Fig 3. (cont.)

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2020/085736

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/09 C12N15/63 A01K67/027
ADD.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
C12N C40B A01K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>ROBERT C. TRYON ET AL: "Development of translating ribosome affinity purification for zebrafish : Zebrafish Translating Ribosome Purification", GENESIS, vol. 51, no. 3, 26 February 2013 (2013-02-26), pages 187-192, XP055691668, ISSN: 1526-954X, DOI: 10.1002/dvg.22363 abstract; figure 1 p. 187 col. 2, last , p. 188 col. 1, last , p. 188 col. 2, last -p. 189 col. 1, 1st , p. 190 col. 1, 2nd -penultimate , p. 191 col. 1, penultimate , p. 190 col. 1, last -p. 190 col. 2, 1st ----- -/--</p>	1-13

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
 "E" earlier application or patent but published on or after the international filing date
 "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
 "O" document referring to an oral disclosure, use, exhibition or other means
 "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
 "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
 "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
 "&" document member of the same patent family

Date of the actual completion of the international search 8 February 2021	Date of mailing of the international search report 24/02/2021
--	--

Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Landré, Julien
--	--

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2020/085736

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Alexander J. Cammack ET AL: "Abstract", bioRxiv, 1 February 2019 (2019-02-01), XP055683908, DOI: 10.1101/538504 Retrieved from the Internet: URL:https://www.biorxiv.org/content/10.1101/538504v2.full.pdf abstract; figure 1 p. 4, 1st -2nd , p. 6, last , p. 7, penultimate -last , p. 6, last , p. 6, last -p. 8, 1st , p. 16, 1st , p. 16, 1st -----	1-13
Y	WO 2004/009792 A2 (KAMINSKI JOSEPH M [US]; UNIV VANDERBILT [US]) 29 January 2004 (2004-01-29) points 6, 280 -----	1-13
Y	Anonymous: "UPI00046002DE", ³ 3 September 2014 (2014-09-03), XP055691714, Retrieved from the Internet: URL:https://www.uniprot.org/uniparc/UPI000 46002DE [retrieved on 2020-05-05] the whole document -----	1-13
Y	Anonymous: "UPI00000FB076", ¹ 1 May 2000 (2000-05-01), XP055691718, Retrieved from the Internet: URL:https://www.uniprot.org/uniparc/UPI000 00FB076 [retrieved on 2020-05-05] the whole document -----	1-13
Y	EP 2 067 402 A1 (MAX DELBRUECK CENTRUM [DE]) 10 June 2009 (2009-06-10) sequence 8 the whole document -----	1-13
Y	Anonymous: "EM_PAT:HV766732", ³ 30 September 2012 (2012-09-30), XP055691725, Retrieved from the Internet: URL:http://ibis.internal.epo.org/exam/dbfe tch.jsp?id=EM_PAT:HV766732 [retrieved on 2020-05-05] the whole document -----	1-13
	----- -/--	

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2020/085736

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 02/44409 A2 (EUROPEAN MOLECULAR BIOLOGY LAB EMBL [DE] ET AL.) 6 June 2002 (2002-06-06) sequence 2 the whole document -----	1-13

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2020/085736

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2004009792	A2	29-01-2004	AT 494370 T 15-01-2011
		AU 2003268022	A1 09-02-2004
		CA 2533708	A1 29-01-2004
		EP 1546322	A2 29-06-2005
		ES 2358994	T3 17-05-2011
		SI 1546322	T1 31-05-2011
		US 2006210977	A1 21-09-2006
		WO 2004009792	A2 29-01-2004

EP 2067402	A1	10-06-2009	EP 2067402 A1 10-06-2009
		WO 2009071334	A2 11-06-2009

WO 0244409	A2	06-06-2002	AU 3203402 A 11-06-2002
		CA 2430378	A1 06-06-2002
		EP 1348028	A2 01-10-2003
		JP 2004518419	A 24-06-2004
		US 2004161756	A1 19-08-2004
		WO 0244409	A2 06-06-2002
