A novel quantitative fluorescent reporter assay for RAG targets and RAG activity

Inês Trancoso, Marie Bonnet, Rui Gardner, Jorge Carneiro, Vasco M. Barreto, Jocelyne Demengeot* and Leonor M. Sarmento

Instituto Gulbenkian de Ciência, Geiras, Portugal

INTRODUCTION

V(D)J recombination, the somatic rearrangement of variable (V), diversity (D), and joining (J) segments of the antigen receptor genes, is the phenomenon responsible for the very large diversity of the B and T cell antigen receptors in jawed vertebrates (Tonegawa, 1983). The Recombination-Activating Genes (RAG) 1 and 2 form the endonuclease that specifically recognizes recombination signal sequences (RSSs) adjacent to each gene segment and generates DNA double strand breaks (DSBs) (Schatz et al., 1989; Oettinger et al., 1990). In a typical V(D)J reaction, the non-homologous end-joining (NHEJ) machinery subsequently processes these DSBs to produce a functional gene. On one side the signal ends retain the RSSs perfectly joined together, whereas the other breaks, named coding ends, are edited prior to ligation, thus creating the junctional diversity that is the signature of V(D)J recombination (Bassing et al., 2002).

The RSS comprises a heptamer and a non-am er separated by a 12 or 23 nucleotide spacer sequence (Gellert, 2002). The heptamer and nonamer, but most notably, the spacer sequences are considerably degenerated, which favors a wide range of interactions with RAG and the fine-tuning of the rearrangement efficiency that is important for the generation of antigen receptor diversity (Cowell et al., 2004). However, a consequence of such degeneracy is that sequences similar to RSSs are found outside of antigen receptor loci. These are named cryptic RSSs (cRSS) and their targeting by RAG has been associated with tumorigenesis (Marculescu et al., 2006; Schlissel et al., 2006). Thus, V(D)J recombination challenges genome integrity and has to be kept under tight control at multiple regulatory levels.

As the breakthrough of RAG1 cloning relied on a reporter plasmid for V(D)J (Schatz et al., 1989; Oettinger et al., 1990), it is not surprising that a number of additional reporters have been designed over the years. To study RAG1/2 activity in eukaryotic cells and to avoid the intricate multi-step of biochemical or prokaryote-based assays (Hesse et al., 1987), reporters in which the readout is based on gain of fluorescence became popular (Liang et al., 2002; Borghesi et al., 2004; Zheng and Schwarz, 2006; Arnal et al., 2010; Scott et al., 2010; Lutz et al., 2011). While each of these reporters has appealing features, none combines a double-fluorescence readout for the presence of the plasmid and its recombined form, the flexibility for transient or stable assays as an episomal or retroviral substrate, and high resolution in detecting RAG-mediated recombination events. In
this work, we describe a novel retroviral-fluorescent reporter of RAG activity (GFPi) and test its function in vitro and in vivo, in stable and transient recombination assays. This tool is sensitive and practical as it allows rapid and quantitative measurement of RAG activity in assays sensitive to RSS nucleotide composition.

MATERIALS AND METHODS

RAG EXPRESSION CONSTRUCTS

The CMV-RAG, H2k-RAG, and the H2k-RAG2ER_TAM expression vectors are depicted in Figure 1. The CMV-RAG expression plasmids were generated in the pCMVβ vector (Clontech) by inserting an XhoI-NorI fragment containing the first exon of the murine RAG2 gene (RAG2 miniexon) adjacent to the XcmI-XcmI murine RAG1 genomic sequence or a XhoI-AseI fragment containing the RAG2 miniexon adjacent to the SalI-AseI murine RAG2 genomic sequence. In the latter case, the stop codon was regenerated upon blunt-end ligation. RAG gene segments are described in (Barreto et al., 2001) where construction of the H2k-RAG vectors is reported.

The H2k-RAG2ER_TAM plasmid was constructed from the DRII-6 plasmid (Oltz et al., 1993) and the pcDNA3-ER vector (kindly provided by P. Coffer). A first intermediate construction, pKS-RAG2 contained the PstI-XhoI RAG2 miniexon and the RAG2 SalI-NotI fragments inserted into pBluescript II KS(−) (Stratagene). Next, pKS-RAG2DStop, contained a SacI digested PCR fragment of the 3′ region of the RAG2 coding sequence generated with the following primers:

- Forward: 5′-CCCGATGCTATGGGAATTCACTGTACGGAGATGACGCT-3′
- Reverse: 5′-CCCGATGCTATGGGAATTCAGAGTTCACTGGAGATGACGCT-3′

This fragment was ligated into the BamHI site of pBluescript II KS(−) to give pKS-RAG2, which contains the SacI digested RAG2 miniexon. The RAG2 ER_TAM construct (E) was created by inserting the G525R mutant form of the estrogen receptor (ER, slant dashed box) that ends in the stop codon (TGA). The CMV expression cassette (A,B) contains the CMV viral promoter (vertical dashed box), an engineered first exon of the RAG2 gene (RAG2 miniexon; ME, small black box), and the SV40 polyA (vertical dashed box). The H2k expression cassette (C,D) contains the H2k mouse MHC class I promoter (gray box), the RAG2 miniexon (ME, small black box), the 3′ untranslated region of the human β-globin gene, including a splice site and the polyA (vertical dashed box), and the Eµ enhancer (gray oval). Restriction sites are identified and depicted as vertical lines (except for: Ali, AseI; Nl, NotI; Bl, BamHI; and BglII).
Trancoso et al. The GFPi assay scores RAG and RSS primers: forward – 5′-TCAACGGAGCTCAATAACC-3′ overlapping the RAG2 SacI restriction site (underlined); reverse – 5′-TGAGGAGCTTTGCTAAATAGATCTAACGTCTTTGATG-3′ with altered nucleotide positions (italic) creating the restriction sites (underlined) BglII, that disrupts the stop codon, and a second SacI site at the 3′ end. The amplicon replaced the SacI fragment of the pKS-RAG2. The next intermediate construct, pcDNA3-RAG2ERTAM contained a SacII-BglII fragment from pKS-RAG2DStop ligated into the HindIII-BamHI digested pcDNA3-ER vector to generate the in-frame fusion of RAG2 with the G525R mutant form of the hormone binding domain of the estrogen receptor. The final H2k-RAG2ERTAM construct was obtained by inserting a Clal-NotI fragment from pcDNA3-RAG2ERTAM into a modified version of pHsE3′ digested with ClalI-EcoRV.

**GFPi CONSTRUCTIONS**

The GFPi reporters of RAG-mediated inversion were built on the previously described pMigR1, a MSCV-internal ribosomal entry site (IRES)-GFP retroviral vector (Pear et al., 1996).

We first constructed the GFPi-red fluorescent protein (RFP) (Figure 2). In the intermediate construction, MSCV-IRES-RFP, the monomeric RFP (mRFP) from pRSETB-mRFP1 (Campbell et al., 2002) replaces GFP in pMigR. The backbone of pMigR1 released from the IRES-GFP fragment by EcoRI and HindIII digestion was ligated to a pMigR1 EcoRI–NcoI (blunt) fragment containing the IRES sequence together with a pRSETB-mRFP1 BamHI (blunt)–HindIII fragment containing the mRFP sequences. The next cloning step consisted in inserting GFP bordered or not with RSS sequences in 3′ of the LTR and 5′ of

![Diagram](https://www.frontiersin.org)
the IRES. Primers were designed to amplify GFP from pMigR1. The forward primer included 22 nucleotides complementary to the 5′ end of GFP (5′- ATGGTGAGCAAGGCGAGGAGGCG-3′) preceded in 5′ with an overhang (5′-CCACC3-′) to form a kozak sequence (CCACCATGGT), a 23-RSS (see Table 1) and a full Smal (5′-CCCAGGG-3′) restriction site. The reverse primer included 23 nucleotides complementary to the 3′ end of GFP (5′- TTACTTGTACAGCTCGTCCATGC-3′) followed by an overhang containing a 12-RSS and a full XhoI (5′-CTCGAG-3′) restriction site. The XhoI-Smal digested PCR product was inserted into the XhoI-HpaI digested MSCV-IRES-RFP vector, placing the GFP sequence in an inverted orientation relative to the 5′LTR.

GFPi-Cerulean Fluorescent Protein (CFP) (Figure 6), a second version of the GFPi reporter, bears the CFP instead of the mRFP and was optimized for medium throughput cloning of RSSs. The CFP sequence was amplified from pBS10 (Rizzo et al., 2004) using a 3′ primer containing a SalI restriction site. The PCR product (SalI-blunt) replaced the Ncol(blunt)-Ncol fragment containing mRFP in MSCV-IRES-RFP. The GFP inverted sequence was cloned as described above but using XhoI and HpaI restriction sites in the 5′ and 3′ primers, respectively, without added RSS. This strategy allowed insertion of oriented sequences in 5′ or 3′ of GFP by ligation to the GFPi-CFP vector submitted to either a BglII and XhoI or HpaI and EcoRI restriction site. Each inserted sequence was made of a pair of single-strand oligonucleotides (forward and reverse sequences) containing the RSS under test, three additional guanosines 5′ of the heptamer to improve RAG-mediated DNA DSB (Yu and Lieber, 1999) flanked by a BglII and XhoI or HpaI and EcoRI restriction sites (Table 2). Upon phosphorylation of each strand separately (1 µg), annealing was allowed along a temperature gradient from 80°C to RT.

Plasmid DNA was prepared using a Miniprep column (Qiagen #12125) and proper inserts confirmed by sequencing. T4 polynucleotide kinase, T4 DNA ligase and Poly Ethylen Glycol were from Fermentas, restriction enzymes from New England Biolabs and oligonucleotides from Sigma-Aldrich.

Recombination signal sequences and primer sequences used to generate each GFPi variant are detailed in Tables 1 and 2 for GFPi-mRFP and GFPi-CFP, respectively.

**CELL CULTURE**

Human embryonic kidney 293T cells were cultured in DMEM medium (Invitrogen) supplemented with 10% Fetal Calf Serum (FCS), 100 U/mL penicillin and streptomycin (Invitrogen), at a temperature of 37°C and 5% CO₂ conditions and re-seeded at low density every 2–3 days. Reh, NALM-6, SUP-T1, Jurkat, HL-60, and K-562 cell lines were cultured in RPMI Glutamax medium (Invitrogen) supplemented with 10% FCS, 100 U/mL penicillin and streptomycin, and 0.55 mM β-mercaptoethanol, at 37°C and 5% CO₂ and kept at a density of 0.25 × 10⁶ cells/mL. Cultures enriched in B cell progenitors were established as described in Carvalho et al. (2001) from bone marrow cell suspensions prepared by flushing femurs and tibias with a 23-gage needle and grown at the density of 1 × 10⁶ cells/mL in Optitmem supplemented with 10% FCS, 100 U/mL penicillin and streptomycin, 0.55 mM β-mercaptoethanol, and 10 ng/mL murine recombinant IL-7 (PeproTech).

### Table 1 | RSSs tested with the GFPi-mRFP variants.

| Designation | Origin | Heptamer | Spacer | Nonamer |
|-------------|--------|----------|--------|---------|
| Con23ᵃ | IgKJ1ᵇ,c,d | CACAGTG | GTAGTACTCCACAGTCGGCTGT | AAAAAAAAA |
| ConPi | Syntheticᵃ | CACAGTG | TTGGACACCACTCGGAGGCTGT | AAAAAAAAA |
| MI | Syntheticᵃ | CACAGTG | TTGCAACCACTACGTGGCTGT | AAAAAAAAA |
| MI-conPi(4) | Syntheticᵃ | CACAGTG | TTGCAACCACTACGTGGCTGT | AAAAAAAAA |
| MI-conPi(14,15) | Syntheticᵃ | CACAGTG | TTGCAACCACTACGTGGCTGT | AAAAAAAAA |
| MI-conPi(19,20) | Syntheticᵃ | CACAGTG | TTGCAACCACTACGTGGCTGT | AAAAAAAAA |
| MI-CAGnon | lgHV1-63*02ᵇ,c,d | CACAGTG | TTGCAACCACTACGTGGCTGT | CAGAAACC |
| Con12ᵇ⁰ | lgKV1-57*02ᵇ,c,d | CACAGTG | CACAGCTG | AAAAAAAAA |
| Jk8-2 | Jk2-2ᵇ,c,d | CACAGTG | GTGGAAATCGTG | GCAACC |
| LMO2 | cRSS²,g | acaaca – CACAGTA¹ | TTGTCTTACCA | GCAATATT |
| SCL | cRSS²,h | acaaca – CACAGCC¹ | TCCCGCATTCCT | GTATATTCG |

ᵃCon23 was paired with each 12-RSS variants.
ᵇCon12 was paired with each 23-RSS variants.
ᶜMus musculus.
ᵈHomo sapiens.
ᵉInternational ImMunoGeneTics information system (http://www.imgt.org).
⁰Cowell et al. (2004).
¹Raghavan et al. (2001),
²Small italic letters are 6 adjacent nucleotides in 5′ of the putative cRSS as found at the locus.
Table 2 | Oligonucleotides used to construct the GFPi-CFP variants.

| Designation | Origin | Extension\(^a\) | Heptamer | Sequence spacer | Nonamer | Extension\(^a\) |
|-------------|--------|-----------------|----------|-----------------|---------|----------------|
| Con23       | lgKJ1\(^b\) | 5′-AACGGG | CACAGTG | GTAAGTCCACTGTCGTGCTG | AAAAAACC | 5′-G3′ |
| 3′          | 3′-TTGCC | GTGCAC | CATCTAGTG | GGCCACTACGACGACCA | TGGTTTGGG | CTAA-5′ |
| Ml-conP(19,20) | Synthetic\(^c\) | 5′-AACGGG | CACAGTG | TTGGCACCAACATCCGAGCTG | AAAAAACC | 5′-G3′ |
| 3′          | 3′-TTGCC | GTGCAC | AAGGTTGTGTTAGACTGCTG | AAAAAACC | 5′-G3′ |
| Con12       | lgKV13-5′-02\(^b\) | 5′-GATCTCGGG | CACAGTG | CTACAGACTGGA | AAAAAACC | 5′-G3′ |
| 5′          | 5′-AGCCC | GTGCAC | GATGTCGACCT | TGGTTTGGG | GAGCT-5′ |
| Wu-Con12    | Synthetic\(^d\) | 5′-AACGGG | CACAGTG | ATACAGACTCTTA | AAAAAACC | 5′-G3′ |
| 3′          | 3′-TTGCC | GTGCAC | TATGTCGAGAT | TGGTTTGGG | CTAA-5′ |
| LMO2        | cRSS\(^e\) | 5′-AACGGG | CACAGTA | TTGGCTTTACCA | GAAATTATT | 5′-G3′ |
| 3′          | 3′-TTGCC | GTGCAT | AACAGAATTGGG | CGTATTAA | CTAA-5′ |
| 3′Dd3       | lgKJ1\(^e\) | 5′-GATCTCGGG | CACAGTG | GTAGTCCACTGTCGTG | AAAAAACC | 5′-G3′ |
| 5′          | 5′-AGCCC | GTGCAC | CATCTAGGTCGACGACGACCA | TGGTTTGGG | GAGCT-5′ |

*The marks 3 or 5 indicate the position of the respective RSS in GFPi-CFP, relative to the inverted GFP sequence.

1International ImMunoGeneTics information system (http://www.imgt.org).
2Cowell et al. (2004).
3Ramsden et al. (1994).
4Dik et al. (2007).
5Extensions are cloning site + GGG when linked to the heptamer or cloning site only when linked to the non-amer.

GFPI BASED IN VITRO RECOMBINATION ASSAY

293T were seeded at a density of 0.5 × 10^6 cells per 6-well plate well in 2.5 mL of medium, co-transfected 24 h after by replacing 600 μL of medium by a transfection mix containing 10 μL of Lipofectamin2000 (Invitrogen), 600 μL of Serum-free Opti-mem and a total of 10 μg of plasmid DNA: 5 μg of GFPI and 5 μg of mock plasmid (negative control); or 2.5 μg of H2k-RAG1 and 2.5 μg of H2k-RAG2; or 2.5 μg of H2k-RAG1 and 2.5 μg of H2k-RAG2ER\(^{\text{TAM}}\), or equimolar amounts of CMV-RAG1 (1.6 μg) and CMV-RAG2 (1.4 μg), unless otherwise indicated (Figures 2 and 3). After 16 h incubation, cells were washed, incubated for 48 h with fresh medium and harvested for flow cytometry analysis of GFP and RFP or GFP fluorescence. In assays carried out with H2k-RAG2ER\(^{\text{TAM}}\), medium was supplemented either with vehicle (2:10000 ethanol, Sigma) or 200 nM 4-hydroxytamoxifen (4-OHT, Sigma).

FLOW CYTOMETRY ANALYSES

For the GFPI-mRFP assays, flow cytometry data were acquired with a MoFlo (Dako Cytomation Beckman Coulter). RFP was excited with a 561-nm CrystaLaser GCL-050-561 50 mW DPSS laser coupled to fiber optics (38 mW output) and emission detected using a 630/75-nm bandpass filter on channel 7 (FL7). GFP was excited with a 488-nm Coherent Sapphire 488-200 CDRH (140 mW output) and emission detected using a 530/40-nm filter pass on channel 1 (FL1). In some cases, experiments were reproduced with data acquired with a BD FACSAria III: RFP was measured on the PE-Texas Red/mCherry channel, excited with a Yellow Green 561 nm Laser and emission detected with a 610/20-nm filter pass; GFP was measured on the FITC channel, excited with a 488-nm Blue Laser and emission detected with a 530/30-nm filter pass.

For the GFPI-CFP assay, data were acquired on a CyAn ADP (Beckman Coulter), a BD FACSAria III or on a LSR Fortessa (BD Bioscience), using a 96-well-plate auto sampler. GFP was excited and detected using the same wavelengths as described above for the MoFlo when using the CyAn ADP or as described for the FACSAria III when using the LSR Fortessa analyzer. The CFP was either excited by a 442-nm laser using a 450/50-nm filter for detection or by a 442-nm Blue-Violet laser and measured using a 470/20-nm filter pass (LSR Fortessa) which considerably increased the mean fluorescence intensity of the CFP. Data were collected with the following programs: FACS Diva (BD Bioscience) for the LSR Fortessa and FACSAria III, Summit (Beckman Coulter) for the Cyan ADP and the MoFlo. In all cases, data analysis was performed using the FlowJo software (Tree Star Inc.).

PCR AND SEQUENCE ANALYSES OF RECOMBINED GFPI

Plasmid DNA was recovered using the Qiagen Spin Miniprep Kit (Qiagen) from GFP\(^{\text{+}}\)/RFP\(^{\text{+}}\) or GFP\(^{\text{−}}\)/RFP\(^{\text{+}}\) cell populations sorted from 293T cells previously co-transfected with GFPI and mock DNA or GFPI and CMV-RAG plasmids, respectively. PCR reactions were performed using the indicated amounts of recovered plasmids as template and the following primers (depicted in Figure 2): forward – 5′-AGCCCTTTTGTACACCTTAAG-3′ and reverse – 5′-GTTGTACTCCAGCTTGCC-3′ to amplify the coding joint (CJ) region of rearranged GFPI (GFPI-RR) plasmids; forward – 5′-CATCTTCCTTAAGGAGCAGG-3′ and reverse – 5′-CCGCGAACTAGCATTAG-3′ to amplify the signal joints (SJ) of GFPI-RR plasmids; and, forward – 5′-AGCCCTTTTGTACACCTTAAG-3′ and reverse – 5′-CATCTTCCTTAAGGAGCAGG-3′ to detect unarranged GFPI (GFPI-UNR). To sequence CJ and SJ, the respective PCR products were separated in an agarose gel and fragments from the expected
FIGURE 3 | GFPi predictably quantifies RAG activity. The 12- and 23-consensus RSS GFPi was co-transfected with serial 2-fold titration of CMV-RAG1/2 constructs (from 1.6/1.4 to 0.2/0.175 µg, final dilution 1/16th of the initial plasmid DNA amount). (A) Representative plots of each RAG expression condition. (B) Recombination efficiency defined by % of GFP+ cells inside an RFP+ gate (test – background) with means and standard deviations of at least three replicates. (C) Relationship between the number of recombined plasmids r in cells containing n total plasmids (left) and of the corresponding GFP and RFP values (right). The theoretical model assumes r is Poisson distributed with mean ρn, where ρ is the recombination coefficient. Left panel: the dots represent two random realizations of the model for ρ = 0.1 (black) and ρ = 0.001 (gray), while the lines represent the mean values. Right panel: the dots represent the logarithm of the RFP (x) and GFP (y) signals corresponding to the number of plasmids in the left panel, assuming that autofluorescence, amplification, variation and measurement noise result in bivariate normal-distributed around log-mean values (curves). (D) Estimation of apparent recombination coefficient ρa in GFPi assays by obtaining the frequency of cells with zero recombination events in binned RFP data (left and center). Left panel: experimental bivariate GFP vs. RFP log-intensity measurements in cells transfected with GFPi alone (without RAG) were fitted by the model log(y) = log(x) + βx, by least square minimization and minimization of residuals mean and trend. The dashed curve represents the best fit and the full black curve represents the best fit added by four times standard deviation of the residuals. Right panel: experimental bivariate GFP vs. RFP log-intensity measurements in cells co-transfected with RAG were binned according RFP log-intensity and the frequency of GFP+ cells in each bin scored (i.e., with GFP values below the black curve). (E) Fitting the model P(r = 0) = exp(−ρ(1 + βx)) to the experimental data. The frequency of GFP− cells was plotted vs. the mean RFP log-intensity per bin for the cells with serial dilutions of CMV-RAG1/2 constructs (A), and fitted by the model. (F) The relative and apparent recombination coefficients (respectively ρ and ρa) in RAG titration experiments are proportional to the expected RAG activity. The estimates of ρa, obtained by model fitting (as in [D]) were normalized such that the relative recombination coefficient ρ at the highest titer is 1. Both sets of value are plotted as a function of the CMV-RAG1/2 titers, ρa = 0.0007 + 0.028(CMV-R1R2), ρ = 0.02 + 0.93(CMV-R1R2). (G) The frequency of GFP+ cells inside the RFP+ gate (as in [B]) is proportional to the logarithm of the relative recombination coefficient ρ.
size extracted using the Gel Extraction Kit (Zymo), cloned into the pGEM-T Easy Vector (Promega) and sequenced with the T7 primer (5′-TAATACGACTCACTATAGG-3′) and the SP6 primer (5′-ATTAGTTACACTATAG-3′).

**VIRAL PRODUCTION AND TRANSDUCTION**

Viral production and transduction were carried out as described in (Sarmento et al., 2005). Briefly, MSCV-IRES-RFP or GFPi were co-transfected into 293T cells using a calcium phosphate precipitation method together with the amphotropic packaging plasmid pKat and the pCMV-VSV-G plasmid encoding the vesicular stomatitis virus G-glycoprotein. Supernatants containing pseudo-typed retrovirus were collected at 48 and 72 h post-transfection and passed through a 0.45-µm filter. Transduction of Reh, NALM-6, SUP-T1, Jurkat, HL-60, and K-562 cell lines as well as mouse bone marrow cells were collected at 1 week in the presence of IL-7 was carried out with 1–2 × 10⁶ cells resuspended in 500 µL of viral supernatant and 500 µL of medium supplemented with Polybrene (Sigma) to a final concentration of 4 µg/mL, followed by spinoculation of 90 min, 37°C at 2200 rpm and incubation for the indicated time.

**BONE MARROW TRANSDUCTION AND TRANSPLANTATION**

Retroviral transduction of mouse bone marrow cells and transfer into lethally irradiated recipients was adapted from (Pear et al., 1996). Bone marrow cells were collected from 6- to 12-week-old C57BL/6 or Rag2 knock-out mice 4 days after intravenous administration of 250 mg/kg of fluorouracil (5-FU, Mayne Pharma). The cells were cultured overnight in IMDM supplemented with 15% FCS, 100 U/mL of penicillin and streptomycin in the presence of the following cytokines (PeproTech): 10 ng/mL IL-3, 5 ng/mL IL-6, and 100 ng/mL SCF. The cells were then washed, resuspended in a mixture with half volume of viral supernatant and half volume of culture medium, supplemented with the same cytokine cocktail and polybrene (4 µg/mL, Sigma), and centrifuged at 2200 rpm, 37°C for 90 min. A second round of spinoculation was performed the following day. After washing with PBS, at least 5 × 10⁶ cells were injected intravenously into lethally irradiated (9 Gy) C57BL/6 recipients. Mice were bled 7 weeks post-transfer and peripheral blood cells were analyzed by flow cytometry. All experiments were performed in accordance with the guidelines for the care and use of animals under an animal protocol approved by the Instituto Gulbenkian de Ciência Animal Care and Use Committee.

**MATHEMATICAL MODELING AND STATISTICAL ANALYSES**

The frequency of GFP⁺ cells is related to the intensity of GFP and RFP signal in each cell using a mathematical that assumes multiple reporter plasmids. We define the recombination coefficient $\rho$ as the probability that each plasmid is independently recombined in the time of the assay. Considering that this probability is relatively low, we assume that the number of recombined plasmids $r$ in a cell containing $n$ total plasmids is approximately Poisson distributed with mean $\rho n$ (illustrated in Figure 3C-left). We are particularly interested in the probability that none of the $n$ plasmids is recombined which is $P(r=0|n) = \exp(-\rho n)$. The expected intensity of the default reporter signal (i.e., GFP) in cells containing $n$ total reporter plasmids is $x_0 + m_x n$, where $x_0$ is the mean basal signal and $m_x$ is the mean signal due to a single plasmid. Likewise, the expected intensity of the recombination-associated signal (i.e., GFP) in cells containing $r$ recombined reporter plasmids is $y_0 + m_r r$, where $y_0$ is the mean basal signal and $m_r$ is the mean signal due to a single recombined plasmid. In practice, random variation expression of two genes of the reporter plasmid and measurement noise leads to a distribution of the measured signals around these expected values (illustrated in Figure 3C-right). From the linear relationship between the signal $x$ and the number of total plasmids one obtains $n = (x - x_0)/m_x$, which allows us to obtain an expression for the probability that the cells did not recombine any of its $n$ reporter plasmids as a function of the measured $x$-signal intensity: $P(r=0|n) = \exp(-\rho(x - x_0)/m_x) = \exp(-\rho x + \rho x_0/m_x)$. The apparent recombination coefficient $\rho_a$ is estimated by fitting the expression to empirical data of the relative frequency of cells without recombined plasmids, denoted $F_0(x)$, as a function of the $x$-signal intensity, i.e., the relative frequency of GFP⁺ cells as a function of intensity of RFP signal (illustrated in Figure 3F).

The data processing, model fitting and estimation were performed in R. Briefly, bivariate (RFP signal, GFP signal) data were exported from FlowJo as scaled values after appropriate gating and subsequently imported into the statistical software R. The function $\log_{10}(y) = \log_{10}(x_0 + b_{RFP} + b_{GFP})$ is fitted to data from samples without nominal RAG activity by minimizing least-squares and ensuring that the residuals have zero mean and no trends, and the standard deviation of the residuals $S_0$ computed. The relative frequency of cells without recombined plasmids in any sample of interest is obtained by binning on the $x$-axis, and computing within each bin the frequency of cells fulfilling the condition $\log_{10}(GFP) \leq \log_{10}(RFP) + b_{GFP} + kS_0$, where $k$ is an appropriate constant. This procedure avoids problems with color compensation and works directly with uncompensated data. The recombination coefficient associated to each sample is obtained by fitting the model $P(r=0) = \exp(-\rho_a(x - x_0))$ to the binned frequencies by non-linear least square. Analysis of simulated data indicated that 11 bins lead to robust estimates and that more accurate and precise estimates are obtained if bins with less than 5% of all the gated events are neglected. Furthermore, $k$ is fixed such that the binned frequencies in samples without RAG are over 99.9% (typically in the range 3–5). This tool is being further developed for adaptation to a user-friendly platform and is available upon request.
The frequency of GFP\(^+\) inside the RFP\(^+\) gate, which is used as an empirical proxy of the RE throughout this article, is proportional to the logarithm of the recombination coefficient (Figure 3G).

**RESULTS**

**GFPI IS A RAG-DEPENDENT RECOMBINATION REPORTER**

We have engineered the reporter for RAG1/2 activity (GFPI) in the MSCV-IRES-RFP retroviral vector backbone that allows the detection of the substrate, revealed by the IRES-driven expression of the RFP. Whenever RAG-dependent recombination events occur, the 12-RSS and 23-RSS sequences are targeted by RAG, leading to the inversion of the Green Fluorescent Protein (GFP) coding sequence and therefore rendering coupled RFP and GFP expression (Figure 2A).

To assess the functionality of our reporter, 293T cells were transfected with GFPI containing consensus 12 and 23-RSS either alone or together with equimolar amounts of CMV-RAG1 and 2 (Figure 2B). The transfection efficiency was routinely \( \geq 50\% \), as revealed by the percentage of RFP\(^+\) cells. In absence of RAG the background GFP signal was always \(<1\%\) while in presence of RAG a RFP\(^+\) GFP\(^+\) cell population was readily detectable. The frequency of transfected cells that underwent recombination is determined by the percentage of GFP\(^+\) cells inside an RFP\(^+\) gate, to which value one can subtract the background defined in absence of RAG.

We next confirmed that the GFP\(^+\) cells carried RAG-dependent rearranged GFPI (Figure 2C). Double-positive RFP\(^+\)GFP\(^+\) cells from cells co-transfected with GFPI and RAG were sort purified, the plasmid DNA recovered and tested for the presence of recombination-specific amplicons by semi-quantitative PCR. Plasmid DNA extracted from RFP\(^+\) cells from cultures transfected with GFPI only served as control. Amplification of signal and CJs resulting from GFPI inversion was readily detectable in the plasmid preparations recovered from RFP\(^+\) GFP\(^+\) cells but not in control RFP\(^+\) cells. As each cell received several copies of the reporter upon transfection, GFPI-UNR amplicons were detectable in both cell populations. The rearranged GFPI was further confirmed to display signature of RAG-dependent recombination by sequencing the PCR products (Tables 3 and 4). As expected, CJs presented end-processing by nucleotide excision and addition as well as palindromic sequences (P nucleotides) at the position of the repaired junction; N nucleotides were not observed nor expected as TdT is not expressed in non-lymphoid cells (Komori et al., 1993).

The processing of the coding joints occasionally affected the GFP sequence, removing one or two nucleotides, an imperfection we corrected in the subsequent version of the GFPI vector (see below and Figure 7). Nevertheless, as upon transfection each cell bears several copies of the reporter upon transfection, GFPI-UNR amplicons were detectable in both cell populations. The rearranged GFPI was further confirmed to display signature of RAG-dependent recombination by sequencing the PCR products (Tables 3 and 4). As expected, CJs presented end-processing by nucleotide excision and addition as well as palindromic sequences (P nucleotides) at the position of the repaired junction; N nucleotides were not observed nor expected as TdT is not expressed in non-lymphoid cells (Komori et al., 1993).

The processing of the coding joints occasionally affected the GFP sequence, removing one or two nucleotides, an imperfection we corrected in the subsequent version of the GFPI vector (see below and Figure 7). Nevertheless, as upon transfection each cell bears several copies of the reporter upon transfection, GFPI-UNR amplicons were detectable in both cell populations. The rearranged GFPI was further confirmed to display signature of RAG-dependent recombination by sequencing the PCR products (Tables 3 and 4). As expected, CJs presented end-processing by nucleotide excision and addition as well as palindromic sequences (P nucleotides) at the position of the repaired junction; N nucleotides were not observed nor expected as TdT is not expressed in non-lymphoid cells (Komori et al., 1993).

**GFPI ASSAY RESPONDS TO RAG ACTIVITY TITRATION IN A PREDICTIVE MANNER**

We next assessed the quantitative response of the GFPI assay to controlled variations in recombination activity. To this end, we titrated down CMV-RAG1/2 plasmids in the transfection mixture by serial twofold dilutions. As shown in Figure 3A, GFPI assay displays a clear dose-dependent response. From the highest to the lowest concentration of RAG expression vectors, the frequency of GFP\(^+\) cells measured by FACS progressively decreases from 60\% to 13\%. The low standard deviation we obtained across repeats and across experiments illustrate that variation in the amount of effectively transfected DNA is negligible. In a first approach we quantified the recombination activity in our assay by simply monitoring the percentage of GFP\(^+\) cells detected in the RFP\(^+\) population. This empirical RE score was calculated by subtracting the background values obtained in control cells transfected with GFPI in absence of RAG (Figure 3B).

However, it was conspicuous that the center of mass of the GFP\(^+\) cells progressively shifts to lower GFP values and higher RFP intensities, as the expected RAG activity decreases. To understand how both the frequency of GFP\(^+\) cells and the GFP intensity depend quantitatively on the underlying RE in the assay we developed a mathematical model. The model describes the number of recombinated targets per cell as a function of the probability of recombinating targets during the assay, that we call recombination coefficient \( \rho \). We assume that the number of recombinated targets is Poisson distributed with mean equal to the product of \( \rho \) by number of targets per cell. Two random realizations of the model for two values of the recombination coefficient (\( \rho = 0.1 \) and \( \rho = 0.01 \)) are depicted in Figure 3C (left), illustrating the discreteness of the number of plasmids around the mean values depicted by the trend lines. The RFP and GFP signals measured by flow cytometry are random continuous variables proportional to the total and recombinated plasmid numbers, respectively, which masks the underlying discreteness as illustrated in Figure 3C (right). In this model, as in the real data, when the recombination coefficient decreases the center of mass of the GFP\(^+\) cells in the bivariate space progressively shifts to lower GFP values and higher RFP values.

The key question then becomes how to infer the recombination coefficient from the experimental bivariate data obtained by flow cytometry? The model suggests two potential approaches. It would be tempting to use the asymptotic linear relationship for large values of reporter per cell, which is predicted in the model and observed in the data in the case of high recombination efficiencies (Figure 3A CMV-R1R2 high titers). However, simulations of the model (not shown) indicate that for low recombination efficiencies this strategy gives estimates that are not accurate. A second alternative to infer the recombination coefficient is to consider the probability that a cell with a given RFP signal intensity makes no recombination during the time of the assay. According to the model this probability is given by \( \exp(-\rho(x - x_0)) \), where RFP intensity measured in the cell, \( x_0 \) is the mean RFP autofluorescence, and \( \rho \) is an apparent recombination coefficient that is proportional to the true underlying recombination coefficient \( \rho \), with the proportionality constant being the average RFP signal obtained from a single epitope. By taking ratios between two values of \( \rho \) obtained in identical experimental settings the proportionality constant cancels and one can obtain the relative ratio between the true \( \rho \) values. Simulations of the model (not shown) indicated that accurate estimate of \( \rho \) and relative \( \rho \) values can be obtained by fitting the above probability as a function of RFP...
Table 3 | Sequences of rearranged GFPi-mRFP Con12/Con23 coding joints.

| Clones | pMSCV5 | BglI | Xhol | Insertions | 3' – Coding end |
|--------|--------|------|------|------------|----------------|
| Germline | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | GCACC | ATGGTGAGCC |
| CJ1    | CTTCCTAGGCGCGCGGAATT | AGATCT | CTC- | - | -TGTTGAGCC |
| CJ2    | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGA- | G | -CC | ATGGTGAGCC |
| CJ3    | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | CT | -CC | ATGGTGAGCC |
| CJ4    | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | CT | -CC | ATGGTGAGCC |
| CJ5    | CTTCCTAGGCGCGCGGAATT | AGATCT | CTC- | - | -CC | ATGGTGAGCC |
| CJ6    | CTTCCTAGGCGCGCGGAATT | AGATCT | CTC- | - | G | ATGGTGAGG |
| CJ7    | CTTCCTAGGCGCGCGGAATT | AGATCT | CTC- | - | - | ATGGTGAGC |
| CJ8    | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | GC | - | ATGGTGAGG |
| CJ9    | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | CTC | - | ATGGTGAGC |
| CJ10   | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | CT | - | ATGGTGAGC |
| CJ11   | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | CTC | - | ATGGTGAGC |
| CJ12   | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | CT | - | ATGGTGAGC |
| CJ13   | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | CT | - | ATGGTGAGC |
| CJ14   | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | CT | - | ATGGTGAGC |
| CJ15   | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | C | - | ATGGTGAGC |
| CJ16   | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | CT | - | ATGGTGAGC |
| CJ17   | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | CT | - | ATGGTGAGC |
| CJ18   | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | CT | - | ATGGTGAGC |
| CJ19   | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | CT | - | ATGGTGAGC |
| CJ20   | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | CT | - | ATGGTGAGC |
| CJ21   | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | CT | - | ATGGTGAGC |
| CJ22   | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | CT | - | ATGGTGAGC |
| CJ23   | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | CT | - | ATGGTGAGC |
| CJ24   | CTTCCTAGGCGCGCGGAATT | AGATCT | CTCGAG | CT | - | ATGGTGAGC |

Deleted nucleotides are represented by dashes and potential P elements marked in italics.

Table 4 | Sequences of rearranged GFPi-mRFP Con12/Con23 SJ.

| Clones | 12-RSS | 23-RSS |
|--------|--------|--------|
| Nonamer | Spacer | Heptamer | Heptamer | Spacer | Nonamer |
| Germline | GTTTTTGT | TCCAGCTCTGTAG | CACTGGT | CACAGTG | CTACAGCTCAGCTGCTACTGGA | AAAAAAC |
| SJ1    | GTTTTTGT | TCCAGCTCTGTAG | CACTGGT | CACAGTG | CTACAGCTCAGCTGCTACTGGA | AAAAAAC |
| SJ2    | GTTTTTGT | TCCAGCTCTGTAG | CACTGGT | CACAGTG | CTACAGCTCAGCTGCTACTGGA | AAAAAAC |
| SJ3    | GTTTTTGT | TCCAGCTCTGTAG | CACTGGT | CACAGTG | CTACAGCTCAGCTGCTACTGGA | AAAAAAC |
| SJ4    | GTTTTTGT | TCCAGCTCTGTAG | CACTGGT | CACAGTG | CTACAGCTCAGCTGCTACTGGA | AAAAAAC |
| SJ5    | GTTTTTGT | TCCAGCTCTGTAG | CACTGGT | CACAGTG | CTACAGCTCAGCTGCTACTGGA | AAAAAAC |
| SJ6    | GTTTTTGT | TCCAGCTCTGTAG | CACTGGT | CACAGTG | CTACAGCTCAGCTGCTACTGGA | AAAAAAC |
| SJ7    | GTTTTTGT | TCCAGCTCTGTAG | CACTGGT | CACAGTG | CTACAGCTCAGCTGCTACTGGA | AAAAAAC |
| SJ8    | GTTTTTGT | TCCAGCTCTGTAG | CACTGGT | CACAGTG | CTACAGCTCAGCTGCTACTGGA | AAAAAAC |
| SJ9    | GTTTTTGT | TCCAGCTCTGTAG | CACTGGT | CACAGTG | CTACAGCTCAGCTGCTACTGGA | AAAAAAC |
| SJ10   | GTTTTTGT | TCCAGCTCTGTAG | CACTGGT | CACAGTG | CTACAGCTCAGCTGCTACTGGA | AAAAAAC |
| SJ11   | GTTTTTGT | TCCAGCTCTGTAG | CACTGGT | CACAGTG | CTACAGCTCAGCTGCTACTGGA | AAAAAAC |
| SJ12   | GTTTTTGT | TCCAGCTCTGTAG | CACTGGT | CACAGTG | CTACAGCTCAGCTGCTACTGGA | AAAAAAC |
| SJ13   | GTTTTTGT | TCCAGCTCTGTAG | CACTGGT | CACAGTG | CTACAGCTCAGCTGCTACTGGA | AAAAAAC |
| SJ14   | GTTTTTGT | TCCAGCTCTGTAG | CACTGGT | CACAGTG | CTACAGCTCAGCTGCTACTGGA | AAAAAAC |
| SJ15   | GTTTTTGT | TCCAGCTCTGTAG | CACTGGT | CACAGTG | CTACAGCTCAGCTGCTACTGGA | AAAAAAC |
values to the relative frequencies of GFP negative cells scored at different log-intensities of RFP signal.

We used this method to infer the apparent and relative recombination coefficients underlying the data obtained for serial dilutions of CMV-RAG1/2 plasmid in Figure 3A. The following procedure was used to compute the frequencies of GFP negative cells for different values of RFP intensities. First, we fitted a linear model to the bivariate data obtained in cells not transfected with CMV-RAG1/2 plasmid (dashed curve in the plots Figure 3D), ensuring that the GFP log-intensity residuals of the fitting (i.e., the differences between observed GFP log-intensity and the expected values) had approximately null values of mean and regression coefficient against the RFP log-intensity (not shown). We then defined a cutoff curve by adding four to five standard deviations of the residuals to the fitted line (black curve in the plots Figure 3D). The number of standard deviations added to the mean value was set to ensure that the frequency of GFP negative cells (i.e., cells with GFP below the cutoff curve) measured in cells without RAG activity is unitary with a precision <10^-4. For each cell population transfected with the different Rag plasmids titer we binned the data according to the RFP log-intensity and computed the frequencies in each bin (illustrated for the unitary CMV-R1/R2 titer in Figure 3D-right). Using the GFP negative frequencies per RFP signal intensity thus obtained for each data set (Figure 3E) we estimated the values of apparent RE (ρa) by non-linear least square fitting of the theoretical probability (Figure 3F). As expected from mass action, the relative recombination coefficient ρ (ρa normalized by the mean value obtained for the CMV-R1/R2 higher titer) in cells transfected with CMV-R1/R2 plasmid increases linearly with the plasmids titer, and the regression line has a regression coefficient close to the unit and predicts approximately zero recombination at zero titer (Figure 3F). This means that the recombination coefficient estimated from the GFPi assay is an accurate quantitative measurement. Furthermore, the frequency of GFP^+ cells within the RFP^- gate obtained directly from gating the flow cytometry data (as in Figure 3B) scales linearly with the logarithm of the relative recombination coefficient ρ (Figure 3G). In turn, this result indicates that this simple frequency, determined by classical FACS analysis, can be used as a quantitative proxy of the RE.

**GFPi QUANTIFIES RAG TRANSCRIPTION AND NUCLEAR TRANSLLOCATION**

We next tested whether the GFPi assay detects quantitative differences in efficiency of RAG1/2 transcription or RAG2 nuclear translocation (Figure 4) Variation in transcription efficiency was achieved by transfection of RAG1/2 expression vectors bearing different promoters, namely H2k or CMV (see Figure 1). Variation in nuclear translocation efficiency was achieved by transfection of a construct that contains the RAG2 coding sequence fused to the ligand-binding domain of the estrogen receptor (see Figure 1). The gene is under the H2k promoter and the fusion protein translocates efficiently to the nucleus in presence of an estrogen analog. This construct was used to generate transgenic mice and confirmed to be able to complement the RAG2 knock-out lymphocyte deficiency in vivo specifically in the presence of the estrogen analog, 4-OHT (Sarmiento and Bonnet, in preparation). The 293T cells were transfected with GFPi together with mock plasmid, H2k-RAG1/2, H2k-RAG1, and H2k-RAG2ER or else CMV-RAG1/2 (Figure 4A). The frequency of GFP^+ cells inside the transfected RFP^- subset were dependent on the strength of the promoter driving RAG expression (3.5 ± 1.1% under H2k vs. 62.5 ± 2.2% under CMV) and of the nuclear translocation of RAG2 (0.9 ± 0.1% upon vehicle alone vs. 7.3 ± 0.8% upon 4-OHT induction) (Figure 4B). The estimated recombination coefficient ρ spans about four orders of magnitude in these different setting (Figure 4C). As an example of an application of the GFPi assay, this experiment indicated that the RAG2ER construct present some degree of leakiness when transfected into untreated 293T. The flow cytometry profiles of the mock (no RAG) and the test (H2k-R1R2ER no 4-OHT) are different with noticeable GFP^+ signal in the latter, absent in the former. This difference was not quantifiable by simple percentage analysis but revealed by a relative recombination coefficient ρ of 2.3 × 10^-4 ± 0.5 × 10^-4, close but above the minimum value detectable in the controls (5 × 10^-5). This case also illustrates that when RAG is limiting, only cells with high RFP expression levels, hence with high copy number of GFPi, rearrange enough GFPi to emit a detectable signal. Monitoring GFP^+ frequency in the last percentiles of RFP intensity also enhances the sensitivity of the analysis, although this approach is more easily confounded by variation in the flow cytometry analyzer settings (not shown).

Finally, we evaluated by western blot analysis that RAG1/2 expression was about 10-fold increased when driven by CMV instead of H2k, after equimolar transfection (data not shown). Analysis of the data obtained with the GFPi assay reported a 15-fold difference in the recombination coefficient ρ between H2k and CMV-driven RAG. Together, these results demonstrate that GFPi, when used as an episomal substrate, provides a faithful quantitative assay to detect recombination driven by ectopic RAG1/2 activity.

**GFPi FUNCTIONS AS AN INTEGRATED SUBSTRATE EX Vivo AND IN Vivo**

We next tested whether integrated GFPi substrates delivered through retroviral infection would specifically recombine in a RAG-dependent manner (Figure 5). We first selected five hematopoietic leukemia cell lines for which RAG1/2 activity has been previously quantified using a classical episomal substrate and readout of recombination by differential antibiotic resistance of subsequently transfected bacteria (Gauss et al., 1998). In this early work the Reh and NALM-6, two B cell acute lymphoid leukemia phenotypically resembling a pre-B developmental stage, displayed efficiency of recombination evaluated as 21.6 and 1.8% respectively; the SUP-T1 non-Hodgkin’s lymphoma, double-positive for CD4 and CD8, scored as 0.09% while the acute myelogenous leukemia HL-60 and the chronic myelogenous leukemia K-562 scored at 0.008 and 0.02% respectively (Gauss et al., 1998). We also tested the Jurkat acute T cell leukemia cells that express a mature TCR and low levels of Rag mRNAs (Roose et al., 2003). These tumor cells were infected with control MSCV-ires-RFP or GFPi retrovirus (Figure 5A). All four lymphoid cell lines (Reh, NALM-6, Jurkat, and SUP-T1) but not the myeloid tumors presented a RFP^+ GFP^+ population when infected with GFPi, confirming the ability of the GFPi retrovirus to detect endogenous RAG1/2 activity. Defining the RE by simple percentage analysis for each cell type
allowed their ranking for RAG activity, in concordance with that established earlier, although we did not detect the residual RAG1/2 activity reported to occur in the leukemias of myeloid lineage HL-60 and K-562 (Gauss et al., 1998). Each cell line presents specific size, volume, granularity, and consequently autofluorescence. Moreover each cell line may differently express the DNA binding factors allowing MSCV driven gene expression. With these uncontrolled variations, a formal quantitative assessment of the specific RAG activity with GFPi is unwarranted.

We next assessed whether GFPi reveals RAG activity in primary cells \textit{ex vivo}. Mouse bone marrow cells from WT and RAG2\(^{-/-}\) animals grown in the presence of IL-7 to support the proliferation and survival of B cell progenitors were infected with control MSCV-IRES-RFP or GFPi retrovirus (Figure 5B). FACS analysis revealed the appearance of a double-positive RFP/GFP population exclusively in the GFPi-infected WT cells.

To test the applicability of the GFPi reporter \textit{in vivo}, bone marrow progenitor cells isolated from WT and RAG2\(^{-/-}\) mice infected with GFPi were transferred into lethally irradiated WT recipient mice (Figure 5C). RFP positive cells were readily detected 7 weeks post-transplant in the peripheral blood of all mice and GFP\(^{+}\) RFP\(^{+}\) cells were detected solely in mice reconstituted with WT GFPi-infected WT cells, restricted to the side-scatter low lymphocyte-enriched subpopulation with a frequency of approximately 50% within this subpopulation. Analysis of peripheral blood cells from WT-reconstituted mice stained for T cell, B cell, and myeloid markers confirmed that GFP positive cells were CD3\(^{+}\) or CD19\(^{+}\) (Mac-1-negative) lymphocytes. Overall these results demonstrate that GFPi is a bona fide RAG1/2 reporter substrate in mice and humans, \textit{ex vivo} and \textit{in vivo}.

\textbf{GFPi IS A FAITHFUL CLASSIFIER OF RSS}

We next determined whether the GFPi reporter was able to classify RSSs according to their RE \textit{in vitro}, similarly to the pJH290 system (Lewis et al., 1988). Six variants of the GFPi reporter were generated all containing the same 12-RSS Con12 paired with specific 23-RSS: ConPI, MI, MI-conPI(4), MI-conPI(14,15), MI-conPI(19,20), and MI-CAGnon, each differing in the 23-spacer and/or the nonamer sequence (Cowell et al., 2004), as detailed in Table 1. When cells were co-transfected with H2k-RAG1/2, the GFPi assay discriminated each 23-RSS along a range of GFP\(^{+}\) RFP\(^{+}\) cell frequency from 0.5 to 6% (Figure 6A). Hence, the assay was sensitive to nucleotide changes in the spacer sequence and to the inhibitory effect of the specific CAGnon nonamer (MI vs. MI-CAGnon). Our ranking of these 23-RSS was overall consistent with previous results making use of the p290T reporter (Cowell et al., 2004), with the exception of the GFPi variants carrying the MI and MI-conPI(4) RSSs. As the 23-RSS containing the MI spacer
FIGURE 5 | GFPi retrovirus is a RAG1/2-specific reporter substrate ex vivo and in vivo. (A) Human lymphoid leukemic cell lines of B (Reh, NALM-6) and T (Jurkat, SUP-T1) cell origin as well as myeloid leukemia and erythroleukemia cell lines (HL-60 and K-562, respectively) were transduced with MSCV-IRES-RFP (vRFP) or vGFPi retrovirus. Frequency of cells that underwent RAG recombination was measured by FACS 1 week post-transduction. (B) Bone marrow cells of wild-type (WT, left) or RAG2 knock-out (R2−/−, right) C57BL/6 mice cultured in the presence of IL-7 and either not infected (ni) or infected with MSCV-IRES-RFP (vRFP) or vGFPi reporter (vGFPi) virus were analyzed by FACS for the expression of RFP and GFP to detect RAG-mediated recombination 1 week post-transduction. The plots are representative of three independent experiments. (C) Lethally irradiated C57BL/6 mice were reconstituted with BM cells isolated from 5-FU-treated R2−/− or WT C57BL/6 mice after transduction with the GFPi or the RFP retrovirus. Shown are representative contour plots (n = 3 for each type of chimeras) of peripheral blood cells analyzed by FACS 7 weeks post-transfer. A first gate, not shown, identified the mononuclear cell population according to the forward and side-scatter physical parameters. Cells were then analyzed for RFP and GFP expression, gated on RFP+ cells (blue gate) and reanalyzed for SSC and GFP. The SSChigh subset is enriched in myeloid cells, while the SSClow population are enriched in lymphoid cells. For WT cells, RFP+ cells were also analyzed for the distribution of lymphocytes (CD3+ or CD19+) by co-staining in the same channel (PE) vs. macrophages/myeloid cells (MAC1/CD11c+). Analysis of GFPbrightRFP+ (purple gate) for the same lineage markers documents rearrangement in lymphocytes.
was described as presenting the highest *in silico* score, according to the RSS Information Content (RIC score), and the highest experimental RE amongst all 23-RSSs tested (Cowell et al., 2004), we were particularly surprised by its low score in our assay. As the CMV-RAG GFPi system appeared to be ideal to study RSSs with low RE, we also tested MI, Con23, and MI-CAGnon in these conditions (Figure 6B) and confirmed that MI has low RE when compared to the classical consensus sequence. The ranking of this set of 23-RSS either by frequency of rearranged cells or by the coefficient of recombination ρ was reproducible across independent experiments (Figure 6C).

We next tested the capacity of the assay to detect 12-RSS of very low functionality. These were paired with 23-RSS Con23 and tested in conditions of high levels of RAG expression (Figure 6D). The frequency of GFP+RFP+ cells were of 2.3 ± 0.02% for the mouse TCR-Jβ2-2 12-RSS, a sequence previously reported to be a poor RSS when compared to a large set of V(D)J associated 12-RSS (Cowell et al., 2003). Putative cryptic 12-RSS (cRSS) have been described in the LMO2 and the SCL genes and proposed to be involved in oncogenic translocations or deletions found in some T-ALL. We could not detect rearrangement when scoring the SCL sequence by simple frequency of GFP+ cells. However the recombination coefficient ρ was low but higher than the respective background (SCL construct, no RAG). This result is in agreement with a previous study (Raghavan et al., 2001). The LMO2 sequence tested positive and scored similarly to the Jβ2-2 sequence. Together, these results confirm that the GFPi reporter assay is suitable to assess RSS and cRSS functions.

**GFPi ADAPTATION TO LARGE-SCALE CANDIDATE RSS CLONING**

The results above indicating that GFPi offers a fast and reliable readout of RAG activity prompted us to modify the original GFPi-mRFP reporter to ease the preparations of variants. We generated GFPi-CFP (Figure 7A), in which the indicator of transfection efficiency is provided by expression of CFP, offering an alternative color for eventual analysis in combination with other fluorescent reporters. To improve the efficiency of RSS insertions, GFPi-CFP was engineered to contain a pair of dissimilar cohesive-end cloning sites in 5′ and 3′ of GFP. Sequences under test were ordered as sense and antisense oligonucleotides bordered by the complementary cloning site ends. Annealing efficiency was optimal upon a temperature gradient thus allowing direct cloning into the digested and dephosphorylated GFPi-CFP. To further improve the GFPi efficiency each (putative) RSS sequences contained and additional GGG in 5′ of the heptamer (Yu and Lieber, 1999). Together, these modifications resulted in an insertion of nine nucleotides between the RSS in 5′ of the inverted GFP and the Kozak sequence (Table 2), dramatically reducing the possibility that nucleotide excision during the processing of the coding ends would affect the GFP sequence, as we had observed with GFPi-RFP (Table 3).

To validate this new tool, we tested the RE of constructs carrying various pairs of RSS when co-transfected with CMV-RAG (Figure 7B). We first compared the 23-RSS con23 and MI-conPI(19,20), each positioned in 3′ of the inverted GFP and paired with the 12-RSS con12. Similarly to the results obtained with GFPi-mRFP (Figure 6A), MI-conPI(19,20) performed better than Con23 (RE 47.7 ± 1.7 vs. RE 38.4 ± 1.4) in the GFPi-CFP assay. We next compared Con12, a sequence widely used as a consensus 12-RSS but actually a physiological RSS found at the Kappa locus, with Wu-Con12, a real consensus 12-RSS defined by analysis of a large number of physiological V(D)J associated RSS (Kamshden et al., 1994). Strikingly, when paired with Con23, Wu-Con12 performed much better than Con12 (% of GFP+ cells 65.3 ± 1.6 vs. 38.4 ± 1.4), confirming that it is a better-defined consensus 12-RSS. Finally, we tested the LMO2 12-cRSS suspected to mediate translocation to the TCRβ locus (Dik et al., 2007). This sequence, when paired with one of its natural partner in malignant translocations, the 23-RSS in 5′ of D3S, performed remarkably well, with a frequency of GFP+ cells of 9.7 ± 0.5, a result consistent with previous studies (Raghavan et al., 2001). With these analyses, we conclude that the GFPi assay, now also available as GFPi-CFP will be a suitable tool to test (putative) RSS properties on a medium to large scale.

**DISCUSSION**

Studies on V(D)J recombination and RAG-mediated genomic instability have relied on tools that provide information on the targeting of candidate substrates by RAG. RAG-mediated recombination reporters are included in those tools as they are key for further characterization of the regulation levels underlying the recruitment of a RSS or a cRSS. In this work, we introduce a reporter assay that combines a direct measurement of transfection/transduction efficiency and of RAG activity without the need for further procedures and we validated it as a tool for quantitative analyses.

Our fast method of readout, relying on a fluorescence-based reporter system to measure RAG activity is not novel. Recently, Scott et al. (2010) have generated the first mutually exclusive dual-fluorescence reporter that accounts a transfection efficiency marker as control but lacks the retroviral structure that enables the stable integration of the reporter. Accounting for differences in both methodologies, we observed that GFPi performs at one-log greater efficiency when tested with the same 12 and 23-RSS sequences. Another construct similar to GFPi, pMX-RSS-GFP-ires-Cd4, is a retroviral-based inversion reporter; that has been used *ex vivo* in primary mouse lymphoid cells, in studies of RAG mutant activities and regulation of recombination impacting on allelic exclusion (Liang et al., 2002; Lutz et al., 2011). When integrated, this reporter reaches recombination efficiencies similar to the ones of GFPi in the same context and with the same RSSs. Yet, contrarily to GFPi, this tool requires indirect detection of CD4 expression by antibody staining. Zheng and Schwarz (2006) have described the pE19HKN, an exclusively episomal GFP-fluorescence inversion reporter in which GFP expression is under the control of the EF-1 alpha promoter while the transfection efficiency is assessed by monitoring expression of an independently CMV-driven hemagglutinin gene, by antibody staining. Contrary to GFPi, transcription of each marker in pE19HKN is controlled independently, which may lower the accuracy of the inferred recombination efficiencies. When transfecting 293 cells with RAG in this assay, recombination frequencies were of 0.6%. Other recombination reporters lack a second marker for transfection control. A mouse transgenic for the VEX-GFP recombination substrate was generated for *in vivo* detection of RAG activity.
FIGURE 6 | The GFPi assay efficiently ranks nucleotide sequences for their RSS functionality. GFPi variants carrying different 12 or 23-RSSs were tested in 293T cells transfected alone (mock) or co-transfected with either H2k-RAG1/2 (A) or CMV-RAG1/2 (B, D) GFPi 23-RSS variants Con23, ConPI, MI-conPI(4), MI-conPI(14,15), MI-conPI(19,20), MI, and MI-CAGnon were described previously (Cowell et al., 2004) and their nucleotide sequences are listed in Table 1. (A-B) Representative FACS analysis gated on RFP+ cells. Number are% of GFP+ cells C Recombination efficiency defined as (%GFP+ in test – (%GFP+ in respective control) (left and middle panel) At least six replicates from at least two independent experiments. Right panel shows the relative recombination coefficient ρ normalized to GFPi-Con. For MI-CAGnon, the less efficient RSS, ρ in the mock was <5.10^-5 D) FACS analysis of 293T cells transfected with CMV-RAG1/2 and GFPi variants carrying 12-RSS of poor efficiencies paired with Con23. Jβ2-2 is found at the TCRβ locus, LMO2 and SCL are putative cRSS. Mock is Jβ2-2 transfected alone. The frequency of GFP+RFP+ cells in an RFP+ gate is indicated in the upper part of the plot and the recombination coefficient ρ relative to GFPi-Con in A, for which ρ = 1.0±0.2) in the bottom part, in blue. Standard deviation of three replicates and their significance vs. the mocks. (**p < 0.05, ***p < 0.001, ns – not significant).
FIGURE 7 | In vitro recombination assay using the GFPi-CFP variant. (A) Linear representation of the GFPi-CFP reporter. (B) The recombination of different RSS pairs was determined upon 293T cells transfection with each GFPi-CFP variants together with the CMV-RAG1/2 expressing plasmids. The 23-RSS Con23 and Mi-conPI(19,20), as well as the 12-RSS Con12 and LMO2 were as in Figure 5. Wu-Con12 is a consensus 12-RSS defined by bio-informatics (Ramsden et al., 1994). The 3′Dδ3 is a physiological partner of LMO2 described in Dik et al. (2007). The specific pairs of RSS tested are indicated above the corresponding FACS plots and the sequences listed in Table 2. The recombination efficiency defined as (% GFP+ cells in test - % GFP+ cells in control) is indicated in the upper part of each plot. The relative recombination coefficient ρ, when con12/con23 is 1, is indicated in the bottom part of each plot, in blue. Four replicates of two independent experiments.

This system served as a RAG lineage tracer, as the RAG-dependent inversion of VEX would render cells VEX+. However, in this case, the system did not allow discriminating truly non-recombining cells from those carrying the silenced transgene. Arnal et al. have used several versions of RAG-mediated deletional reporters that conditioned the expression of GFP. These were used in vitro in the context of NHEJ and homologous recombination repair (Arnal et al., 2010). In both cases, the performance of these reporters cannot be directly compared to the one of GFPi. When analyzing all RAG-mediated reporter systems, none of the former combined all features presented by GFPi, namely: (1) a double-fluorescence system; (2) its usage in transient or stable assays as an episomal or retroviral substrate; (3) high resolution; (4) a tool here thoroughly validated for RAG and RSS studies.

We have shown here that GFPi can be used as a semi-quantitative assay by measuring the frequency of GFP+ cells as a proxy of the RE. This straightforward measurement can be obtained directly from the standard flow cytometry software. Nevertheless, the dynamic range of this measurement is limited, and as an alternative we proposed a rigorous quantitative method to estimate the relative recombination coefficient (i.e., the probability that a target episome is recombined) in the GFPi assay. The method provides estimates that span about four orders of magnitude and seem to be both accurate and precise. The variance within the coefficients estimates in independent transfection replicates is remarkably small when compared to the range of the estimates. It is noteworthy that this method is formally analogous to the classical limiting dilution analysis (LDA). The RFP signal and the apparent recombination coefficient (ρa) in each cell is mathematically analogous to the cellular input and expected responder frequency in LDA (Carneiro et al., 2009). In this formal analogy each cell in the GFPi assay represents an independent replicate culture in the LDA, which allows to compute accurate and precise frequencies from very large number of “replicates” (>1000), which are obviously prohibitive when performing limiting dilutions. For this reason our method is effectively better with frequencies of negative cells that are in the range 30–100% in contrast with LDA, which operates better in the range 0–50%. This provides an intuition for why our method seems to be able to resolve recombination coefficients across several orders of magnitude with relatively high precision.

As a tool that scores RAG activity, GFPi will be useful for the study of mutant RAGs, and can serve as a functional diagnosis
of Rag mutation in patients with primary immunodeficiencies, as previously done with other reporters (Schwarz et al., 1996; De Ravin et al., 2010). In agreement with this proposition, we demonstrated that GFPi reveals both mouse and human RAG activities. The former was evidenced in transfected fibroblasts, in ex vivo cultures of pre-B cells and during hematopoiesis in vivo while the latter was tested in human cell lines of lymphoid origin. As a tool to detect endogenous RAG activity in vivo, GFPi has the potential to serve in lineage-tracing and cell fate studies in hematopoiesis. The high resolution and double reporting features of GFPi also make it suitable to address in vivo the lingering questions of physiological RAG expression and activity in non-lymphoid cells (Chun et al., 1991) and to further elucidate the conditions that may promote RAG re-expression in mature lymphocytes (Hikida et al., 1996) and in lymphoid or non-lymphoid tumors (Gashaw et al., 2005; Marculescu et al., 2006; Schlissel et al., 2006; Mcintyre et al., 2007).

We have validated GFPi as a tool able to reveal RSSs and cRSSs, scoring each according to their RE in vitro. We have used GFPi to score a set of synthetic 23-RSSs and detected the effect of single nucleotide changes in the spacer sequence as well as action of exogenous RAG to run our GFPi assay. Moreover, we used a fibroblast cell line transiently expressing the Rag1 and Rag2 genes in the genome by statistical repair pathways. Nucleic Acids Res. 38, 2944–2954. Barreto, V., Marques, R., and Demenget, J. (2001). Early death and severe lymphopenia caused by ubiquitous expression of the RAG1 and Rag2 genes in mice. Eur. J. Immunol. 31, 3763–3772.

Bassing, C. H., Swat, W., and Alt, F. W. (2002). The mechanism and regulation of chromosomal V(D)J recombination. Cell 109, 545–555.

Borghesi, L., Hsu, L. Y., Miller, J. P., Anderson, M., Herzenberg, L., Schlissel, M. S., et al. (2004). B lineage-specific regulation of V(D)J recombinease activity is established in common lymphoid progenitors. J. Exp. Med. 199, 491–502.

Campbell, R. E., Tzur, O., Palmer, A. E., Steinbach, P. A., Baird, G. S., Zacharias, D. A., et al. (2002). A monomeric red fluorescent protein. Proc. Natl. Acad. Sci. U.S.A. 99, 7877–7882.

Carneiro, J., Duarte, L., and Padovan, E. (2009). Limiting dilution analysis of antigen-specific T cells. Methods Mol. Biol. 514, 95–105.

Carvalho, T. L., Mota-Santos, T., Campos, A., Demengeot, J., and Vieira, P. (2001). Arrested B lymphopoiesis and persistence of activated B cells in adult interleukin 7(-/-) mice. J. Exp. Med. 194, 1141–1150.

Chun, J. J., Schatz, D. G., Oettinger, M. A., Jaenisch, R., and Baltimore, D. (1991). The recombination activating gene-1 (RAG-1) transcript is present in the murine central nervous system. Cell 64, 189–200.

Cowell, L. G., Davila, M., Kepler, T. B., and Kelsoe, G. (2002). Identification and utilization of arbitrary correlations in models of recombination signal sequences. Genome Biol. 3:research0072. doi:10.1186/gb-2002-3-12-research0072

Cowell, L. G., Davila, M., Ramsden, D., and Kelsoe, G. (2004). Computational tools for understanding sequence variability in recombination signals. Immunol. Rev. 200, 57–69.

Cowell, L. G., Davila, M., Yang, K., Kepler, T. B., and Kelsoe, G. (2003). Prospective estimation of recombination signal efficiency and identification of functional cryptic signals in the genome by statistical modeling. J. Exp. Med. 197, 207–220.

De Ravin, S. S., Cowen, E. W., Zarember, K. A., Whiting-Theobald, N. L., Kuhns, D. B., Sandler, N. G., et al. (2010). Hypomorphic Rag mutations can cause destructive midline granulomatous disease. Blood 116, 1263–1271.

Dik, W. A., Nadel, B., Przybylek, G. K., Asnafi, V., Grabarcezyk, P., Navarro, J. M., et al. (2007). Different chromosomal breakpoints impact the level of LMO2 expression in T-ALL. Blood 110, 388–392.

Gashaw, L., Grummer, R., Klein-Hitpass, L., Dushaj, O., Bergmann, M., Brehm, R., et al. (2005). Gene signatures of testicular seminoma with emphasis on expression of its variant gene 4. Cell. Mol. Life Sci. 62, 2359–2368.

Gauss, G. H., Domain, I., Hsieh, C. L., and Lieber, M. R. (1998). V(D)J recombination activity in human hematopoietic cell lines correlates with developmental stage and genome stability. Eur. J. Immunol. 28, 351–358.

Gellert, M. (2002). V(D)J recombination: RAG proteins, repair factors, and regulation. Annu. Rev. Biochem. 71, 101–132.
McIntyre, A., Summersgill, B., Lu, Y. J., Missiaglia, E., Kitazawa, S., Oosterhuis, J. W., et al. (2007). Genomic copy number and expression patterns in testicular germ cell tumours. Br. J. Cancer 97, 1707–1712.

Oettinger, M. A., Schatz, D. G., Gorka, C., and Baltimore, D. (1990). RAG-1 and RAG-2 genes that synergistically recombine V(D)J. Science 248, 1517–1523.

Ohmori, H. (1996). Reexpression of Gokka, C., and Baltimore, D. (1990). RAG-1 and RAG-2, adjacent genes that synergistically recombine V(D)J. recombination. Science 248, 1517–1523.

Oltz, E. M., Ah, F. W., Lin, W. C., Chen, J., Taccioni, G., Desiderio, S., et al. (1993). A V(D)J recombimase-inducible B-cell line: role of transcriptional enhancer elements in directing V(D)J recombination. Mol. Cell. Biol. 13, 6223–6230.

Pear, W. S., Aser, J. C., Scott, M. L., Hasserjian, R. P., Soffer, B., Sklar, J., et al. (1996). Exclusive development of T cell neoplasms in mice transplanted with bone marrow expressing activated Notch alleles. J. Exp. Med. 183, 2283–2291.

Raghavan, S. C., Kirsch, I. R., and Lieber, M. R. (2001). Analysis of the V(D)J recombination efficiency at lymphoid chromosomal translocation breakpoints. J. Biol. Chem. 276, 29126–29133.

Ramsden, D. A., Baetz, K., and Wu, G. E. (1994). Conservation of sequence in recombination signal sequence spacers. Nucleic Acids Res. 22, 1795–1796.

Rizzo, M. A., Springer, G. H., Granada, B., and Pinston, D. W. (2004). An improved cyan fluorescent protein variant useful for FRET. Nat. Biotechnol. 22, 445–449.

Roose, J. P., Diehn, M., Tomlinson, M. G., Lin, J., Alizadeh, A. A., Botstein, D., et al. (2003). T cell receptor-independent basal signaling via Erk and Abl kinases suppresses RAG gene expression. PLoS Biol. 1:E53. doi:10.1371/journal.pbio.0000053

Sarmiento, L. M., Huang, H., Limon, A., Gordon, W., Fernandes, J., Tavares, M. J., et al. (2005). Notch1 modulates timing of G1-S progression by inducing SKP2 transcription and p27kip1 degradation. J. Exp. Med. 202, 157–168.

Schatz, D. G., Oettinger, M. A., and Baltimore, D. (1989). The V(D)J recombination activating gene, RAG-1. Cell 59, 1035–1048.

Schlissel, M. S., Kuffer, C. R., and Curry, J. D. (2006). Leukemia and lymphoma: a cost of doing business for adaptive immunity. Genes Dev. 20, 1539–1544.

Schwarz, K., Gauss, G. H., Ludwig, L., Pannicke, U., Li, Z., Lindner, D., et al. (1996). RAG mutations in human B cell-negative SCID. Science 274, 97–99.

Scott, G. B., de Wynter, E. A., and Cook, G. P. (2010). Detecting variable (V), diversity (D) and joining (J) gene segment recombination using a two-colour fluorescence system. Mol. DNA 1, 9.

Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature 302, 575–581.

Yu, K., and Lieber, M. R. (1999). Mechanistic basis for coding end sequence effects in the initiation of V(D)J recombination. Mol. Cell. Biol. 19, 8094–8102.

Zheng, X., and Schwarz, K. (2006). Making V(D)J rearrangement visible: quantification of recombination efficiency in real time at the single cell level. J. Immunol. Methods 315, 133–143.