The handle http://hdl.handle.net/1887/32970 holds various files of this Leiden University dissertation.

**Author:** Koole, Wouter  
**Title:** Microsatellite and G-quadruplex instability in worm, fish and man  
**Issue Date:** 2015-05-13
Chapter 2

A versatile microsatellite instability reporter system in human cells

Wouter Koole1*, Henning S. Schäfer1, Reuven Agami2, Gijs van Haaften2, Marcel Tijsterman1

*These authors contributed equally to this work
1Department of Toxicogenetics, Leiden University Medical Center, Leiden, The Netherlands
2Division of Gene Regulation, The Netherlands Cancer Institute, Amsterdam, The Netherlands

Adapted from Koole et al. Nucleic Acids Res. 41, e158 (2013)
ABSTRACT

Here, we report the investigation of microsatellite instability (MSI) in human cells with a newly developed reporter system based on fluorescence. We composed a vector into which microsatellites of different lengths and nucleotide composition can be introduced between a functional copy of the fluorescent protein mCherry and an out-of-frame copy of EGFP; \( \text{in vivo} \) frameshifting will lead to EGFP expression, which can be quantified by Fluorescence Activated Cell Sorting (FACS). Via targeted recombineering, single copy reporters were introduced in HEK293 and MCF-7 cells. We found predominantly -1 and +1 base pair frameshifts, the levels of which are kept in tune by mismatch repair. We show that tract length and composition greatly influences MSI. In contrast, a tracts’ potential to form a G-quadruplex structure, its strand orientation or its transcriptional status is not affecting MSI. We further validated the functionality of the reporter system for screening microsatellite mutagenicity of compounds and for identifying modifiers of MSI: using a retroviral miRNA expression library, we identified miR-21, which targets MSH2, as a miRNA that induces MSI when overexpressed. Our data also provide proof of principle for the strategy of combining fluorescent reporters with next generation sequencing technology to identify genetic factors in specific pathways.

INTRODUCTION

The human genome is full of DNA repeats. One abundant class of repeats, making up for \(-3\%\) of the human genome (Lander et al., 2001), are microsatellites, which are often defined as repetitive runs of DNA sequences consisting of 1-8 base pairs (bp) long units (Richard et al., 2008). Soon after their discovery in the early 1980s it became apparent that these tandem repeats are highly polymorphic in length, and have mutation rates even up to \(10^{-7} \) per locus per generation (Ellegren, 2004). It is their repetitive nature that makes microsatellites prone to mutagenesis; due to strand slippage during DNA replication or unequal recombination, microsatellites can expand or contract. Microsatellites can be found all over the genome, present even in protein-coding sequences (Woerner et al., 2009). In fact, 17% of human genes contain tandem repeats in their open reading frames (ORFs) (Gemayel et al., 2010), and microsatellites have been shown to affect biological processes such as chromatin organization, recombination, DNA replication, transcription and translation (reviewed in (Li et al., 2002)). It is therefore of no surprise that microsatellites are thought to play a significant role in evolution, and that many diseases, including several neurodegenerative diseases, and cancer are linked to variations in the length of genomic microsatellites.

The stability of microsatellites is influenced by several factors. An important factor is the status of Mismatch Repair (MMR). This pathway is well-conserved among species and consist of a delicate interplay of many proteins (for a review see for example (Peña-Diaz and Jiricny, 2012) and references therein). In brief, mis-incorporated nucleotides or small insertion-deletions loops are recognized by a heterodimeric protein complex consisting of MSH2 and MSH3 or MSH6. These mutS complexes interact with the mutL proteins MLH1 and PMS2, which are essential for incision and subsequent removal by EXO1 of the newly synthesized DNA. Numerous other proteins (e.g. PCNA, RFC, polymerase-\( \delta \), RPA and DNA ligase I) are required to complete the faithful repair of a mismatch or loop. Another important determinant that affects the stability of microsatellites is the length (the number of repeat-units) of the tract. While a correlation between the length of the microsatellite and the mutation rate has been seen in numerous organisms (Beck et al., 2003; Brinkmann et al., 1998; Brohede et al., 2002; Harr and Schlötterer, 2000; Primmer et al., 1998; Wierdl et al., 1997; Yamada et al., 2002), thus far there is no consensus whether this is a linear, quadratic or exponential relationship (Bhargava and Fuentes, 2009; Brinkmann et al., 1998; Lai and Sun, 2003). Also, the genomic environment of the microsatellite is an important determinant for MSI: ample evidence exists that the locus where the microsatellite is situated is greatly affecting its stability (Harfe and Jinks-Robertson, 2000; Hawk et al., 2005; Kondrashov and Rogozin, 2004; Ma et al., 2012). For example, a recent report showed that the presence of other repeats in close proximity of a microsatellite decreases its stability (Ma et al., 2012). Other factors like nucleotide composition, possible formation of secondary structures such as G-quadruplex structures and
levels of transcription of the locus have also been implicated in the stability of microsatellites (as reviewed in ref. (Bhargava and Fuentes, 2009)).

Many aspects on microsatellite dynamics have been studied in a plethora of organisms. However, several aspects have not been addressed in human cells, despite the notion that microsatellite dynamics clearly vary between organisms (even between humans and chimpanzees) (Webster et al., 2002). To gain full insight into MSI in human cells, we developed an experimental setup that is able to quantify MSI in human cells. We monitor MSI using a modular fluorescent reporter system in combination with FACS. To exclude the influence of the genomic environment, we targeted different microsatellites to the same genomic locus. We addressed the influence of length, orientation, nucleotide composition, secondary structure, the transcriptional status of the locus as well as compound exposure. In addition, we show how this system aids to identify and characterize genetic regulators of MSI by assaying ~450 miRNAs. This methodology can be easily adapted to read out other genome instability phenotypes in mammalian cells to uncover novel regulators in a specific pathway.

MATERIAL AND METHODS

Plasmid construction and sequencing

Standard molecular cloning techniques were used to obtain the constructs described in this manuscript. Briefly, using polymerase chain reaction (PCR), we amplified 3 DNA fragments: mCherry (from plasmid pRSET-B mCherry) without termination codon, flanked by a Nhel and a HindIII restriction-site, a coding stuff fragment of 215 basepairs (bp) flanked by a BamHI and an EcoRI restriction site, EGFP without start-codon flanked by EcoRI and EcoRV restriction sites (from pEGFP-N2, Clontech). Pieces were sequentially cloned into plasmid pcDNA5/FRT/TO (Life Technologies). Microsatellites were subsequently placed into the HindIII and BamHI site using oligo-cloning. All plasmids sequences were checked by Sanger sequencing according to standard procedures, but with the following adjustments: we used a 1:3 mix of ABI Prism dGTP BigDye terminator v3.0 and BigDye terminator v3.0, 200ng plasmid per sequencing reaction, and no more than 25 cycles.

Cell culture

Flp-in T-Rex-293 cells (Life Technologies) were cultured on poly-l-lysine coated surface at 37°C and 5% C0 2 in DMEM (ref 41966, Gibco) supplemented with 10% fetal bovine serum (Bodinco BV) and 1% penicillin-streptomycin. Stable polyclonal surface at 37°C and 5% C0 2 in DMEM (ref 41966, Gibco) supplemented with 10% fetal bovine serum (Bodinco BV) and 1% penicillin-streptomycin. Flp-in T-Rex 293 cell lines were generated via integration of the plasmids using Flp recombinase-mediated DNA recombination according to manufacturer’s protocol. Flp-in T-Rex-293 cells (Life Technologies) were cultured on poly-l-lysine coated surface at 37°C and 5% C0 2 in DMEM (ref 41966, Gibco) supplemented with 10% fetal bovine serum (Bodinco BV) and 1% penicillin-streptomycin. Stable polyclonal surface at 37°C and 5% C0 2 in DMEM (ref 41966, Gibco) supplemented with 10% fetal bovine serum (Bodinco BV) and 1% penicillin-streptomycin. Cells were cultured in the presence of 100µg/ml hygromycin (ref 1084355001, Roche) and 15µg/ml blasticidin (ref A11139-03, Gibco) and 0.1µg/ml doxycycline hyclate (ref D9891, Sigma). MCF-7 cells with a single copy integrated pFRT/lacZeo and the murine ecotrophic receptor were cultured under similar conditions, however selected with hygromycin and neomycin.

Transient knockdown of MSH2

150.000 Cells were seeded per well (12-well plate) per condition. Cells were transfected with a mix of 4ul DharmaFECT1, Sul 20µM ONTARGETplus smartpool MSH2 siRNA, or non-targeting siGENOME Control pool (REF T-2001-03, L-003909-00, D-001206-13-05, Thermo Scientific Dharmacon) and 200ul optimem after 24 hours. 48 hours later, cells were sorted and 25.000 mCherry+EGFP- cells per well were seeded in a 96 well plate. A second round of knockdown was performed 24 hours after and cells were analysed 6 days later.

Fragment analysis

mCherry-EGFP- cells were isolated from HEK 293 cells carrying plm405 (G14-repeat) by FACS. 48 Hours later flow cytometry was used to sort one cell per well in a 96-well plate. Cells were grown to large colonies and checked by eye for expression of mCherry and EGFP. Next, DNA was isolated by NaCl/EtOH precipitation from single cell colonies (mCherry+EGFP+, mCherry+EGFP-, mCherry-EGFP-). A standard PCR was performed using GoTag DNA polymerase (REF M3175, Promega) and oligos CAGTCATAGCCGAATAGCCTCT and 6-FAM-labeled GACCACCTACAAGGCCAGAGA. Samples were run on an ABI 3730 analyser (Applied Biosystems) and analysed with Peak Scanner software v1.0.

Tests for transcription, ICR191 and Phen-DC6

Transcription

Cells were cultured in medium containing 0.1µg/ml doxycycline hyclate and charcoal stripped fetal bovine serum (ref A15-119, PAA laboratories). At day 1 mCherry+EGFP-cells with plm212 (C18-repeat) were sorted and cultured in doxycycline hyclate free medium. At day 3 a similar sort was performed but now 4 cells per well were plated in a 96-well plate with or without doxycycline hyclate. At day 25, when wells were confluent, doxycycline hyclate was added to all plates. Cells were analysed for mCherry and EGFP expression by flow cytometry after 3 days.

Phen-DC6

5000 mCherry+GFP- cells of a polyclonal cell line with plm273 (G11AG11) were plated per well (96 well plate). Cells were transfected with 5µM Phen-DC6 (kindly provided by M.P. Teulade-Fichou, C Guetta and A. Nicolas) or control (DMSO). Cells were quantified by flow cytometry after 6 days.

ICR191

5000 mCherry-GFP- cells (with plm315, G710-repeat) were plated per well (96 well plate). 24 Hours later cells were incubated with various concentrations of ICR 191(ref I3636,
Sigma) dissolved in 0.01M HCL and optimem for two hours. Cells were analysed by flow cytometry after 9 days.

miRNA screen
An MCF-7 reporter cell line was created by single copy integration of plm191 (G23- repeat). Next, a miRNA expression library (miR-Lib) (Voorhoeve et al., 2006) was used to transduce the reporter cell line with ±450 different miR-Vecs by retroviral transduction in a 96-well format in triplicate. Per experiment, cells were drug-selected for 5 days and all stable cell lines with integrated miR-Vecs were combined in one large pool. Next, mCherry+EGFP- cells were selected by flow cytometry and after 21 days mCherry+EGFP+ and mCherry+EGFP- populations were sorted per pool and genomic DNA was isolated (DNeasy kit, Qiagen). The miRNA coding regions from the retroviral (genomically integrated) miR-Vecs were amplified using IlluSeq_IndXX_Mirvec_f and P7_MirVec_r primers, followed by a second amplification step using P5_iillumeseq and P7_MirVec_r primers (sequences can be found in Supplementary Table S1). The resulting library was deep-sequenced on an Illumina Hiseq platform according to manufacturers protocol. The resulting reads were aligned to the miR-lib resulting in 2.5 million aligned reads divided over 6 barcodes (triplicate measurements of two cell populations).

Flow cytometry
Cells were sorted with flow cytometer BD FACSAria III and analysed with flow cytometers BD FacsDiv (BD biosciences) and guava easyCyte HT (Millipore) and their respective software. To set gates we used control polyclonal cell lines that consisted out of solely mCherry+EGFP- or mCherry+EGFP+ cells.

RESULTS AND DISCUSSION

Construction and validation of a fluorescent microsatellite reporter
To study which factors contribute to the stability of microsatellites in human cells, we wished to develop a fast and reliable system that measures MSI independent of its genomic context. Therefore, we decided to make use of FACS enabling measurements of $10^3$ to $10^4$ events per hour and site-specific recombination to target microsatellites at the same specific locus, which ensures the identical genomic environment for each fragile tract, and thus complements and further extends previously described MSI-reporters that require episomal replication or random integration (Cejka et al., 2003; Gasche et al., 2003; Hile et al., 2000).

Accordingly, we created a transgene reporting MSI by expression of the green fluorescent protein EGFP upon a frameshift; we placed the sequence of EGFP downstream of a microsatellite containing ORF such that a -1bp frameshift at the microsatellite leads to in frame EGFP. The upstream ORF encodes functional red fluorescent protein mCherry (illustrated in Figure 1A). mCherry expression visualizes the presence of the reporter as well as the transcriptional activity of the transgene. Unique restriction sites flanking the microsatellite allow for easy insertion of different fragile sites. The sequence of this dual-fluorescent fusion product was cloned into the vector pcDNAs5/FRT/TO (LifeTechnologies) that allows tetracycline-inducible expression under the control of a cytomegalovirus promoter (Andersson et al., 1989; Boshart et al., 1985; Hillen and Berens, 1994; Hillen et al., 1983; Nelson et al., 1987). The vector also contains a single FLp Recombination Target (FRT) site that is required for Flp recombinase-mediated integration of the vector into a genome that contains a single copy integrated FRT site (Craig, 1988; Sauer, 1994). This system allowed us to efficiently create stable cell lines with a single copy integrated reporter containing a microsatellite.

![Figure 1](https://via.placeholder.com/150)

Figure 1 | A fluorescent-based MSI reporter system at a fixed position in the human genome. (A) Schematic representation of the reporter transgene: the coding sequence of EGFP is placed in the +1 frame downstream of 23 guanines (G23) and the coding sequence of mCherry. A -1 frameshift brings the EGFP ORF in frame with the upstream mCherry. As a result, cells will express both markers. TRE = Tetracycline responsive element. (B) A representative image of an MSI event. In cultures, MSI events will manifest as single cells or a small clonal group of cells. Notably, mCherry+EGFP+ cells showed reduced expression levels of mCherry compared to mCherry+EGFP- cells. This image is an overlay of pictures taken in bright-field, green and red channel. (C) Representative FACS plot of a population Flp-In T-Res 293 cells with an integrated copy of the reporter transgene. The vast majority are mCherry+EGFP-, but also mCherry+EGFP+ are seen. (D) Knockdown of MSH2 by siRNA results in increased numbers of mCherry+EGFP+ cells (7 days after pre-sorting mCherry+EGFP- cells). Error bars represent the standard error of the mean (s.e.m.) of three experiments. (E) Exponential relationship between the number of nucleotides in a microsatellite (X-axis) and the incidence of frameshifting, as measured by the percentage of mCherry+EGFP+ cells (Y-axis) 2 days after transduction in a 96-well format in triplicate. Per experiment, cells were drug-used to transduce the reporter cell line with ±450 different miR-Vecs by retroviral (genomically integrated) miR-Vecs were amplified using IlluSeq_IndXX_Mirvec_f and P7_MirVec_r primers, followed by a second amplification step using P5_iillumeseq and P7_MirVec_r primers (sequences can be found in Supplementary Table S1). The resulting library was deep-sequenced on an Illumina Hiseq platform according to manufacturers protocol. The resulting reads were aligned to the miR-lib resulting in 2.5 million aligned reads divided over 6 barcodes (triplicate measurements of two cell populations).
Finally, we tested a reporter containing a microsatellite of 23 guanines in the coding strand (plm191, Figure 1A) in Flp-In T-REX-293 cells that contain a single copy integrated Flp Recombination Target site (LifeTechnologies). As expected, the majority of cells expressed mCherry but not EGFP (mCherry+EGFP-) after integration of the reporter, however a small percentage (±0.2%) also expressed EGFP (mCherry+EGFP+), as visualized by microscopy and FACS (Figure 1B and 1C). As a control we integrated a similar reporter but without a microsatellite (plm184). This cell line showed only mCherry+EGFP- cells (data not shown). Together, this indicated that EGFP expression was the consequence of the presence of the microsatellite. To further substantiate that the EGFP+ cells were the result of MSI, we performed a transient knockdown of the MMR-factor MSH2 by siRNA on pre-sorted mCherry+EGFP- cells. Because the MSH2 protein was reported to be stable (Diouf et al., 2011), two rounds of siRNA-treatment were performed, and FACS-analysis was performed after 7 days of culturing. This resulted in a ±17-fold increase in the number of mCherry+EGFP+ cells, indicative for MSI-dependent expression of EGFP (Figure 1D). Together, these data indicate that EGFP expression is a read-out for MSI. The advantage of the mCherry internal control lies in the fact that by pre-sorting mCherry+EGFP- cells we can i) discard cells that became EGFP positive before an experimental condition was applied, and ii) get rid of cells that lost the reporter or its expression due to silencing.

Exponential correlation between length of a monotract and MSI

To determine the type of correlation between the length of the microsatellite tract and its instability, we made transgenes carrying monotracts of lengths 8, 11, 14, 17, 20 and 23 nucleotides, of all four different nucleotides. On integration in 293-T-REX cells stable polyclonal lines were obtained. Unexpectedly, we found that adenine and thymine monotracts diminished EGFP expression levels (when put in frame as a control), making MSI analysis of these tracts impossible. Therefore, we pursued analysis with the lines containing monotracts of guanines and cytosines. We established populations of 100% mCherry+EGFP- cells (by FACS) and determined the percentages of mCherry+EGFP+ after 48 h of culturing. We started to detect MSI in cell lines containing monotracts of 11bp or longer; in cells with monotracts of 8 nucleotides (either guanine or cytosine) no mCherry+EGFP+ cells were detected, corresponding to a mutation frequency that is < 10⁻⁷. This result may be explained by the inability of a small monotract to induce slippage events: several in vitro and computational studies have argued a threshold for slippage being ~7-9 basepairs (Kelkar et al., 2010; Lai and Sun, 2003; Paoloni-Giacobino et al., 2001; Rose and Falush, 1998; Shinde et al., 2003). Figure 1E indicates that the correlation of MSI and the length of a monotract is best fitted by an exponential trend line; for every extra guanine or cytosine in the monotract the number of mCherry+EGFP+ cells increased with 1.66- and 1.70-fold, respectively (R²= 0.80 and 0.43). We found no significant difference in MSI between cell lines containing similar sized guanine or cytosine monotracts, suggesting no influence of strand orientation, similar to what was reported in yeast (Henderson and Petes, 1992).

Interruption of pure repeats

To explain the distribution and stability of microsatellites across genomes, several groups have postulated mutational models (reviewed in (Bhargava and Fuentes, 2009)). Experimental evidence, as described above, helps to improve these mutational models. Rates for expansions and contractions are obviously important parameters in these models, however another factor of great influence is the effect of point mutations that split up a monotract into two smaller ones. To measure the effect of such an event experimentally, we analysed a reporter (plm273) with an instability locus that is similar to a G23 monotract, apart from having an A instead of a G at position 12, hence comprising two monotracts of 11 Gs (G₁₁AG₁₁). We found that the MSI-rate of this fragile site was > 100-fold lower than that of a G23 monotract (Figure 1E). These results show that a single mutation can lead to stabilization of a repeat with more than two orders of magnitude. Similar observations were found in human mismatch repair-deficient cells (Boyer et al., 2008). Interestingly, the G₁₁AG₁₁ repeat was still 10-fold more unstable than a single G11 repeat, suggesting that two repeats in close proximity are more unstable than the sum of two single G11 repeats.

Contraction exceeds expansion of longer monotracts

During the course of our experiments we noticed the presence of mCherry-EGFP-cells by microscopy and FACS (Figure 1C, lower gate). These cells were resistant to the transgene encoded selection-marker which is located directly upstream of the fluorescent reporter, making silencing or loss of the locus unlikely. Interestingly, we noticed that the fraction of mCherry-EGFP- cells was similar to the fraction of mCherry+EGFP+ cells. We thus wondered whether mCherry-EGFP- cells are the result of +1 frameshift events: although the monotracts were cloned downstream of mCherry to not interfere with its expression, it may be that the -1 frame of GFP encodes an amino acid sequence that interferes with proper mCherry expression. To test this explanation, we constructed transgenes with EGFP in the -1 reading frame. Indeed, cells with integration of such constructs resulted in three populations of cells: the vast majority of cells being mCherry-EGFP-, and two smaller populations of mCherry+EGFP- and mCherry+EGFP+ cells (Figure 2A). To confirm that the latter two populations represented -1bp and +1bp frameshifts, respectively, we performed fragment analysis on these populations. A cell line containing a G14 monotract was used because analysis of long repeats is technically challenging. Indeed, fragment analysis on colonies derived from single cells confirmed that mCherry+EGFP- represented -1 frameshifts, whereas mCherry+EGFP+ represented +1 frameshifts (see Supplementary Figure S1). We thus serendipitously constructed a transgenic reporter that allows the quantification of both contractions and expansions of microsatellites. Subsequently, we used this experimental setup to investigate whether -1 and +1 events are induced with equal frequencies. We isolated mCherry-EGFP- cells of transgenes containing a G₁₄ monotract and analysed the populations 48 hours later. We found 12 times more mCherry-EGFP- cells than mCherry+EGFP+, indicating
that 1 bp contractions are more prevalent than 1 bp expansions. This result supports studies reporting that longer microsatellites are more prone to contracting (Harr and Schlötterer, 2000; Xu et al., 2000).

No role for on-going transcription on MSI levels

Studies in bacteria and yeast have shown that transcription can destabilize simple repetitive DNA sequences (Kiyama and Oishi, 1994; Wierdl et al., 1996). We questioned if and to which extent transcriptional activity affects MSI in human cells by monitoring frameshifting at mononucleotides in cells that were cultured in the presence or absence of doxycycline, a regulator of the Tet repressor that controls the levels of transcription over the transgene (Yao et al., 1998). To avoid transcriptional activity of the transgene at the time of pre-sorting, cells were grown in doxycycline-free medium for three days, resulting in low but still detectable levels of mCherry. Then, four single mCherry-EGFP- cells were seeded and the cultures were grown till confluency, while the transcriptional status of the transgene was visualized by mCherry expression (Figure 3A). To be able to use EGFP as a marker, doxycycline was also added to the non-transcribed conditions three days prior to reading out MSI. In contrast to studies in bacteria and yeast, MSI is not elevated by ongoing transcription in human cells (Figure 3B).

MSI-reporter applicable for testing compounds

One of the applications of the system we developed resides in testing compounds for possible mutagenic characteristics in mammalian cells. As proof of principle we used the compound 6-chloro-9-[3-(2,2'-chloro-ethylamino)propylamino]-2-methoxy-acridine (ICR191), an acridine derivate that is a known inducer of frameshift mutations (Ferguson and Denny, 1990). We exposed mCherry-EGFP- cells with a G<sub>n</sub>-repeat to various concentrations of ICR191 and allowed them to grow for several days, before analysis by flow cytometry. Figure 3C shows a dose-dependent increase of both -1 and +1 frameshifts. In addition, we observed, in line with previous studies (Calos and Miller, 1983; Cariello et al., 1990; Chen et al., 2000; Ferguson and Denny, 1990; Taft et al., 1994), that -1 frameshifts were more frequently induced by ICR191 than +1 frameshifts (53-fold versus 12-fold, respectively).

Quadruplex formation and MSI

Non-B-DNA conformations of microsatellites are thought to facilitate slippage and thus influence the frameshift mutation rate (Bhargava and Fuentes, 2009). One of such non-B-DNA structures is a G-quadruplex structure; a secondary structure that consists of minimally 3 stacked planar arrays of 4 guanines held together by Hoogsteen hydrogen bonding (Bochman et al., 2012). In this study, we have used several microsatellites that have the ability to form a G-quadruplex structure. To address a possible influence of G-quadruplex formation on MSI, we assayed cells containing a G<sub>n</sub>-AG<sub>n</sub>-repeat, a sequence that has the potential to form a G4-quadruplex structure in vivo (Koole et al., unpublished), in the presence and absence of Phex-DC<sub>n</sub> (De Cian et al., 2007) an effective stabilizer of G4-structures in yeast and human cells (Halder et al., 2011; 2012; Lopes et al., 2011; Piazza et al., 2010). We found, however, similar level of MSI in cells treated or mock-treated with to Phex-DC<sub>n</sub> (Figure 3C), suggesting that quadruplex-formation does not stimulate frameshift mutations in mononucleotides.
a miRNA expression library consisting of expression constructs (miR-Vecs) for most human miRNAs (Voorhoeve et al., 2006). Stable cell lines with integrated miR-Vecs were pooled and mCherry+EGFP- cells were selected by flow cytometry to remove cells with mutation events that occurred during the transduction and selection process. After three weeks of cultivating, mCherry+EGFP- and mCherry+EGFP+ cells were sorted and genomic DNA was extracted and subjected to next generation sequencing to determine the relative abundance of each miRNA insert per population. Out of a total of 301 datapoints for individual miR-Vecs (Figure 4B), we selected 17 miRNAs to determine the relative abundance of each miRNA insert per population. Out of these 17 miRNAs, only miR-21 reproducibly induced MSI. Cell growth and plating efficiency was not affected by miR-21 overexpression (data not shown). Several target genes have been described for miR-21, including the mismatch repair gene MSH2 (Valer et al., 2010). The microRNA binding prediction program targetscan predicts a binding site for miR-21 in the 3' UTR of human MSH2 mRNA (Figure 4C). Using immunoblotting experiments, we indeed observed a visible reduction of MSH2 protein levels after miR-21 overexpression (Figure 4D) arguing that MSH2 is a bona fide target of miR-21. To further substantiate this interaction, we cloned the MSH2 3' UTR downstream of the Renilla luciferase gene. Co-transfection of miR-21 with this reporter construct (psiCHECK2-MSH2) in HEK 293 cells (which have low miR-21 levels (Zhu et al., 2008)) led to a ~40% decrease in ren/ffly luciferase activity. We used this assay to screen for miRNAs affecting MSI. MSI was scored using the integrated reporter system in human cells (Figure 4A). The Y-axis represents the Log2 ratio of the relative abundance of a miR-Vec in EGFP+ over EGFP− cells. The X-axis represents the Log2 count per miR-Vec, as quantified by deep sequencing. The V-axis represents the Log2 ratio of the relative abundance of a miR-Vec in EGFP(+) over EGFP(−) cells. (C) Predicted conserved binding site in 3' UTR of MSH2. (D) Immuno-detection of MSH2 protein levels in MCF-7 cells with mCherry+EGFP+ HEK 293 cells carrying a single copy C23 reporter transgene in the absence or presence of doxycycline. Images are overlay of pictures taken in bright-field, green and red channel. (B) Percentage of mCherry+EGFP+ HEK 293 cells carrying a single copy C23 reporter were grown for 6 days in the presence or absence of doxycycline was added to all cells to read out reporter expression. The graph represent the data of ~140 individually-grown populations per condition derived from three independent experiments. Error bars denote the median with interquartile range. Percentage of median is shown. (C) mCherry-GFP- HEK 293 cells carrying a single copy G11 reporter were grown for 6 days in the presence of Phen-DC6 and analysed by flow cytometry. Data represent 32 replicates derived from two independent experiments. Error bars denote s.e.m, n.s= not significant by unpaired two-tailed t-test.
3′UTR of MSH2, we mutated its binding site by site-directed mutagenesis (Figure 4C), resulting in reporter construct psiCHECK2-MSH2-mut. These mutations killed the down-regulatory effect of miR-21 overexpression on luciferase expression (Figure 4E), thus providing proof for specificity in miR-21's genetic interaction with the evolutionary conserved binding site in the 3′ UTR of MSH2. These results also provide proof of principle that our newly developed reporter system can be used to find (novel) bona fide genetic modifiers of MSI.

CONCLUSION

In this study we describe a functional fluorescence-based reporter that reads out MSI in vivo at one specific locus and can be measured by flow cytometry allowing for high-throughput analysis. To prove the functionality of this reporter and meanwhile obtaining biological informative data we tested several repeat-intrinsic and environmental factors that can influence MSI in human cells. In addition, candidate genes can be screened for factors that are potentially implicated in MSI. We combined retroviral transduction of several hundred miRNA-overexpressing constructs with next generation sequencing to identify a regulator of MSI when overexpressed. Although we made use of a miRNA-library and an MSI-specific construct with next generation sequencing to identify a regulator of MSI when overexpressed, we would like to point out that this approach is applicable to any other type of library (e.g. compound or knockdown libraries) and is easily adaptable to readout other potential fragile sequences or other specific pathways when using adjusted fluorescent reporters.

ACKNOWLEDGEMENTS

The authors thank M.P. Teulade-Fichou, C. Guetta and A. Nicolas for providing the compound Phen-DC6.

FUNDING

European Research Council [203379, DSBrepair, to M.T.]; European Commission (DDR response); VENI-NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek) to G.v.H.; Forschungsstipendium der Deutsche Forschungsgemeinschaft to (H.S.S.); Funding for open access charge: European Research Council

Conflict of interest statement. None declared.

REFERENCES

Andersson, S., Davis, D.L., Dahlbäck, H., Jönsson, H., and Russell, D.W. (1989). Cloning, structure, and expression of the mitochondrial cytochrome P-450 steroid 26-hydroxylase, a bile acid biosynthetic enzyme. J. Biol. Chem. 264, 8222–8229.

Beck, N.R., Double, M.C., and Cockburn, A. (2003). Microsatellite evolution at two hypervariable loci revealed by extensive avian pedigrees. Mol. Biol. Evol. 20, 54–61.

Bhargava, A., and Fuentes, F.F. (2009). Mutational Dynamics of Microsatellites. Mol Biotechnol 44, 250–266.

Bochman, M.L., Paeschke, K., and Zakian, V.A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. Nat. Rev. Genet. 1–11.

Boshart, M., Weber, F., Jahn, G., Dorsch-Hälder, K., Fleckenstein, B., and Schaffner, W. (1985). A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. Cell 41, 521–530.

Boyer, J.C., Hawk, J.D., Stefanovic, L., and Farber, R.A. (2008). Sequence-dependent effect of interruptions on microsatellite mutation rate in mismatch repair-deficient human cells. Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis 640, 89–96.

Brinkmann, B., Klinitsch, M., Neuhuber, F., Hühne, J., and Rolf, B. (1998). Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. The American Journal of Human Genetics 62, 1408–1415.

Brohede, J., Primmer, C.R., Møller, A., and Ellegren, H. (2002). Heterogeneity in the rate and pattern of germline mutation at individual microsatellite loci. Nucleic Acids Res. 30, 1997–2003.

Calos, M.P., and Miller, J.H. (1981). Genetic and sequence analysis of frameshift mutations induced by ICR.191. J. Mol. Biol. 153, 39–64.

Carestio, N.F., Koehang, P., Kat, A.G., and Thilly, W.G. (1990). Molecular analysis of complex human cell populations: mutational spectra of MNNG and ICR-191. Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis 231, 165–176.

Cejka, P., Marra, G., Hemmerle, C., Cannavo, E., Storchova, Z., and Jiricny, J. (2003). Differential killing of mismatch repair-deficient and proficient cells towards the therapy of tumors with microsatellite instability. Cancer Res. 63, 8113–8117.

Chen, W.D.W., Estleman, J.R.I., Aminoharai, M.R.M., Ma, A.H.A., Veloso, N.N., Markowitz, S.D.S., Sedwick, W.D.W., and Veigl, M.L.M. (2000).

Cytotoxicity and mutagenicity of frameshift-inducing agent ICR191 in mismatch repair-deficient colon cancer cells. J Natl Cancer Inst 92, 480–485.

Craig, N.L. (1988). The mechanism of conservative site-specific recombination. Annu. Rev. Genet. 22, 77–105.

De Cian, A., Delsomos, E., Mergny, J.L., Teulade-Fichou, M.P., and Monchaud, D. (2007). Highly efficient G-quadruplex recognition by bisquinolinium compounds. J. Am. Chem. Soc. 129, 1856–1857.

Douf, B., Cheng, Q., Krynetskaia, N.F., Yang, W., Cheok, M., Pei, D., Fan, Y., Cheng, C., Krynetsky, E.Y., Geng, H., et al. (2011). Somatic deletions of genes regulating MSI2 protein stability cause DNA mismatch repair deficiency and drug resistance in human leukemia cells. Nature Medicine 17, 1298–1303.

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. Nat. Rev. Genet. 5, 435–445.

Ferguson, J.R., and Denny, W.W. (1990). Framshift mutagenesis by acridines and other reversibly binding DNA ligands. Mutagenesis 5, 529–540.

Gasche, C., Chang, C.L., Nataraajan, L., Goel, A., Rhee, J., Young, D.J., Arnold, C.N., and Boland, C.R. (2003). Identification of frame-shift intermediate mutant cells. Proc. Natl. Acad. Sci. U.S.A. 100, 1914–1919.

Gemayel, R., Vincen, M.D., Legendre, M., and Verstrepen, K.J. (2010). Variable Tandem Repeats Accelerate Evolution of Coding and Regulatory Sequences. Genetics 44, 445–477.

Halder, K., Largy, E., Benzler, M., Teulade-Fichou, M.P., and Hartig, J.S. (2011). Efficient suppression of gene expression by targeting 5′-UTR-based RNA quadruplexes with bisquinolinium compounds. ChemBiochem 12, 1663–1668.

Halder, R., Rau, J.-F., Teulade-Fichou, M.-P., Frickey, T., and Hartig, J.S. (2012). Bisquinolinium compounds induce quadruplex-specific transcriptome changes in HeLa S3 cell lines. BMC Res Notes 5, 138.

Harfe, B.D., and Jinks-Robertson, S. (2000). Sequence composition and context effects on the generation and repair of frameshift intermediates in mononucleotide runs in Saccharomyces cerevisiae. Genetics 156, 571–578.

Harr, B.B., and Schlüttner, C.C. (2000). Long microsatellite alleles in Drosophila melanogaster have a downward mutation bias and short persistence times, which cause their genome-wide underrepresentation. Genetics 155, 1213–1220.
Hawk, J.D., Stefanovic, L., Boyer, J.C., Peters, T.D., and Farber, R.A. (2003). Variation in efficiency of DNA mismatch repair at different sites in the yeast genome. Proc. Natl. Acad. Sci. U.S.A. 102, 8639–8643.

Henderson, S.T., and Peters, T.D. (1992). Instability of simple sequence DNA in Saccharomyces cerevisiae. Mol. Cell. Biol. 12, 2749–2757.

Hile, S.E., Yan, G., and Eckert, K.A. (2000). Somatic mutation rates and specificities at TC/AG and GT/CA microsatellite sequences in nontumorigenic human lymphoblastoid cells. Cancer Res. 60, 1698–1703.

Hillen, W., and Berens, C. (1994). Mechanisms underlyng expression of Ts10 encoded tetracycline resistance. Annu. Rev. Microbiol. 48, 345–369.

Hillen, W., Gatt, C., Autschmid, L., Schollmeier, K., and Meier, I. (1983). Control of expression of the Ts10-encoded tetracycline resistance genes. Equilibrium and kinetic investigation of the regulatory reactions. J. Mol. Biol. 169, 707–721.

Kelkar, Y.D., Strubezowski, N., Hile, S.E., Chiaromonte, F., Eckert, K.A., and Makova, K.D. (2010). What is a microsatellite: a computational and experimental definition based upon repeat computational behavior at A/T and G/C repeats. Genome Biol. Evol. 2, 620–635.

Kiyama, R., and Oshia, M. (1994). Instability of plasmid DNA maintenance caused by transcription of poly(T)-containing sequences in Escherichia coli. Gene 150, 57–61.

Kondrashov, A.S., and Rogozin, I.B. (2004). Context of deletions and insertions in human coding sequences. Hum. Mutat. 23, 177–185.

Lai, Y., and Sun, F. (2003). The relationship between microsatellite slippage mutation rate and the number of repeat units. Mol. Biol. Evol. 20, 2123–2131.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Li, Y.-C., Korol, A.B., Fahima, T., Belles, A., and Nevo, E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol. Ecol. 11, 2453–2465.

Lopes, I., Piazza, A., Bermejo, R., Kriegsman, B., Colosimo, A., Teulade-Fichou, M.-P., Foaani, M., and Nicolas, A. (2011). G-quadruplexes induced instability during leading-strand replication. Embo J 30, 4033–4046.

Ma, X., Rogacheva, M.V., Nishati, K.T., Zanders, S., Bustamante, D.C., and Alani, E. (2012). Mutation Hot Spots in Yeast Caused by Long-Range Clustering of Homopolymeric Sequences. CellReports 1, 36–42.

Nelson, J.A., Reynolds-Kohler, C., and Smith, B.A. (1987). Negative and positive regulation by a short segment in the 5'-flanking region of the human cytomegalovirus major immediate-early gene. Mol. Cell. Biol. 7, 4115–4129.

Paolini-Giacobino, A., Rossier, C., Papasavvas, M.P., and Antonarakis, S.E. (2001). Frequency of replication/transcription errors in (A)n(T)n runs of human genes. Hum. Genet. 109, 40–47.

Peña-Diaz, J., and Jercinovic, J. (2012). Mammalian mismatch repair: error-free or error prone? Trends in Biochemical Sciences 37, 206–214.

Piazza, A., Boule, J.-B., Lopes, J., Ming, K., Largy, E., Teulade-Fichou, M.-P., and Nicolas, A. (2010). Genetic instability triggered by G-quadruplexes interacting Phen-DC compounds in Saccharomyces cerevisiae. Nucleic Acids Res. 38, 4337–4348.

Pruim, C.R., Sano, N., Moller, A.P., and Ellegren, H. (1998). Unraveling the processes of microsatellite evolution through analysis of germ line mutations in barn swallows Hirundo rustica. Mol. Biol. Evol. 15, 1047–1054.

Richard, G.F., Kerrest, A., and Dujon, B. (2008). Comparative Genomics and Molecular Dynamics of DNA Repeats in Eukaryotes. Microbiology and Molecular Biology Reviews 72, 686–727.

Rose, O., and Falush, D. (1998). A threshold size for microsatellite expansion. Mol. Biol. Evol. 15, 613–615.

Sauer, B. (1994). Site-specific recombination: developments and applications. Curr. Opin. Biotechnol. 5, 521–527.

Shinde, D., Lai, Y., Sun, F., and Arrtheim, N. (2003). Tag DNA polymerase slippage mutation rates measured by PCR and quasi-likehood analysis: (CA)GTn and (A)Tn microsatellites. Nucleic Acids Res. 31, 974–988.

Taft, S.A., Liber, H.L., and Skopek, T.R. (1994). Mutational spectrum of ICR-191 at the hprt locus in human lymphoblastoid cells. Environ. Mol. Mutagen. 23, 96–100.

Valeri, N.N., Gasparini, P.P., Braccon, C.C., Paone, A.A., Lott, F.F., Fabbi, M.M., Sumanii, E.M.K., Alder, H.H., Amadori, D.D., Patel, T.T., et al. (2010). MicroRNA-21 induces resistance to 5-fluorouracil by down-regulating human DNA Muts homolog 2 (MMS2). Proc. Natl. Acad. Sci. U.S.A. 107, 21098–21103.

Voorhoeve, P.M.P., le Sage, C.C., Schrier, M.M., Gillis, A.J.M.A., Stoop, H.H., Nagel, R.R., Liu, Y.-P., van Duinse, J.J., Drost, J.J., Grinklepoor, A.A., et al. (2006). A Genetic Screen Implicates miRNA-372 and miRNA-373 As Oncogenes in Testicular Germ Cell Tumors. Cell 124, 13–43.

Webster, M.T., Smith, N.G.C., and Ellegren, H. (2002). Microsatellite evolution inferred from human-chimpanzee genomic sequence alignments. Proc. Natl. Acad. Sci. U.S.A. 99, 8748–8753.

Wierdl, M., Dominka, M., and Peters, T.D. (1997). Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics 146, 769–779.

Wierdl, M., Greene, C.N., Datta, A., Johns-Robertson, S., and Peters, T.D. (1996). Desaturation of simple repetitive DNA sequences by transcription in yeast. Genetics 143, 713–721.

Wloch, S.M., Yuan, Y.P., Benner, A., Korf, S., Knebel Doeberitz, von, M., and Bork, P. (2009). SelfTarbase, a database of human mononucleotide-microsatellite mutations and their potential impact to tumorigenesis and immunology. Nucleic Acids Res. 38, D682–D689.

Xu, X., Peng, M., and Fang, Z. (2000). The direction of microsatellite mutations is dependent upon allele length. Nat. Genet. 24, 396–399.

Yamada, N.A., Smith, G.A., Castro, A.P., Roques, C.N., Boyer, J.C., and Farber, R.A. (2002). Relative rates of insertion and deletion mutations in dinucleotide repeats of various lengths in mismatch repair proficient mouse and mismatch repair deficient human cells. Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis 499, 213–225.

Yasu, F., Svennoji, T., Winkler, T., Liu, M., Eriksson, C., and Eriksson, E. (1998). Tetracycline repressor, tetR, rather than the tetR-mammalian cell transcription factor fusion derivatives, regulates inducible gene expression in mammalian cells. Hum. Gene Ther 9, 1939–1950.

Zhu, S., Wu, H., Wu, F., Nie, D., Sheng, S., and Mo, Y.-Y. (2008). MicroRNA-21 targets tumor suppressor genes in invasion and metastasis. Cell Res. 18, 350–359.
Supplementary Table 1 | Primers used for amplification of miR-Vecs

| Primer | Sequence |
|--------|----------|
| IlluSeq_Ind01_MirVec_f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTaaGCTTGGTACCGAGCTCGGATC |
| IlluSeq_Ind02_MirVec_f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTatGCTTGGTACCGAGCTCGGATC |
| IlluSeq_Ind03_MirVec_f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTagGCTTGGTACCGAGCTCGGATC |
| IlluSeq_Ind04_MirVec_f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTacGCTTGGTACCGAGCTCGGATC |
| IlluSeq_Ind05_MirVec_f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTtaGCTTGGTACCGAGCTCGGATC |
| IlluSeq_Ind06_MirVec_f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTttGCTTGGTACCGAGCTCGGATC |
| IlluSeq_Ind07_MirVec_f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTtgGCTTGGTACCGAGCTCGGATC |
| IlluSeq_Ind08_MirVec_f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTtcGCTTGGTACCGAGCTCGGATC |
| IlluSeq_Ind09_MirVec_f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTgaGCTTGGTACCGAGCTCGGATC |
| IlluSeq_Ind10_MirVec_f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTtgGCTTGGTACCGAGCTCGGATC |
| IlluSeq_Ind11_MirVec_f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTggGCTTGGTACCGAGCTCGGATC |
| IlluSeq_Ind12_MirVec_f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTgcGCTTGGTACCGAGCTCGGATC |
| P5_IlluSeq | AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| P7_MirVec_r | CAAGCAGAAGACGGCATACGAGATCAACCTACAGTTGGGTTCCTTC |
| P7 | CAAGCAGAAGACGGCATACGAGAT |
| P5 | AATGATACGGCGACCACCGAGATCT |
This image shows kidney cells, that contain a reporter construct, resulting in expression of green fluorescent protein upon lengthening of a microsatellite (a repetitive DNA sequence) or expression of red fluorescent protein upon shortening of the microsatellite.