Rapid identification of homologous recombinants and determination of gene copy number with reference/query pyrosequencing (RQPS)

Zhenyi Liu  
*Washington University School of Medicine in St. Louis*

Anna C. Obenauf  
*Medical University of Graz*

Michael R. Speicher  
*Medical University of Graz*

Raphael Kopan  
*Washington University School of Medicine in St. Louis*

Follow this and additional works at: [https://digitalcommons.wustl.edu/open_access_pubs](https://digitalcommons.wustl.edu/open_access_pubs)

**Recommended Citation**
[https://digitalcommons.wustl.edu/open_access_pubs/1940](https://digitalcommons.wustl.edu/open_access_pubs/1940)

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact engeszer@wustl.edu.
Rapid identification of homologous recombinants and determination of gene copy number with reference/query pyrosequencing (RQPS)

Zhenyi Liu, Anna C. Obenauf, Michael R. Speicher, et al.

*Genome Res.* 2009 19: 2081-2089 originally published online October 1, 2009
Access the most recent version at doi:10.1101/gr.093856.109

---

**Supplemental Material**
http://genome.cshlp.org/content/suppl/2009/10/02/gr.093856.109.DC1.html

**Related Content**
A method for rapid, targeted CNV genotyping identifies rare variants associated with neurocognitive disease
Heather C. Mefford, Gregory M. Cooper, Troy Zerr, et al.
*Genome Res.* September, 2009 19: 1579-1585
Sensitive and accurate detection of copy number variants using read depth of coverage
Seungtai Yoon, Zhenyu Xuan, Vladimir Makarov, et al.
*Genome Res.* September, 2009 19: 1586-1592

**References**
This article cites 44 articles, 21 of which can be accessed free at:
http://genome.cshlp.org/content/19/11/2081.full.html#ref-list-1

Articles cited in:
http://genome.cshlp.org/content/19/11/2081.full.html#related-urls

**Creative Commons License**
This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at http://creativecommons.org/licenses/by-nc/3.0/.

**Email Alerting Service**
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

---

To subscribe to *Genome Research* go to:
http://genome.cshlp.org/subscriptions

Copyright © 2009 by Cold Spring Harbor Laboratory Press
Rapid identification of homologous recombinants and determination of gene copy number with reference/query pyrosequencing (RQPS)

Zhenyi Liu, Anna C. Obenauf, Michael R. Speicher, and Raphael Kopan

Department of Developmental Biology, Washington University School of Medicine, St. Louis, Missouri 63110, USA; Division of Dermatology, Washington University School of Medicine, St. Louis, Missouri 63110, USA; Institute of Human Genetics, Medical University of Graz, Harrachgasse 21/8, A-8010 Graz, Austria

Manipulating the mouse genome is a widespread technology with important applications in many biological fields ranging from cancer research to developmental biology. Likewise, correlations between copy number variations (CNVs) and human diseases are emerging. We have developed the reference-query pyrosequencing (RQPS) method, which is based on quantitative pyrosequencing and uniquely designed probes containing single nucleotide variations (SNVs), to offer a simple and affordable genotyping solution capable of identifying homologous recombinants independent of the homology arm size, determining the micro-amplification status of endogenous human loci, and quantifying virus/transgene copy number in experimental or commercial species. In addition, we also present a simple pyrosequencing-based protocol that could be used for the enrichment of homologous recombinant embryonic stem (ES) cells.

[Supplemental material is available online at http://www.genome.org.]
Here we describe a novel approach based on converting these quantitative problems into a determination of single nucleotide variation (SNV) frequencies by pyrosequencing. Pyrosequencing sequences single-strand DNA by coupling nucleotide incorporation onto a sequencing primer with the production of bioluminescence signals, producing readouts called “pyrograms” (Ronaghi et al. 1998) (http://www.pyrosequencing.com/). Each peak in a pyrogram corresponds to a specific nucleotide, and the height is proportional to the amount of nucleotide incorporated during each synthesis step (see Fig. 2; Ronaghi et al. 1998). Pyrosequencing has gained enormous popularity in DNA methylation analysis (Uhlmann et al. 2002; Tost and Gut 2007), microbial identification (Clarke 2005), detection of allelic expression imbalance (Wang and Elbein 2007), and next-generation sequencing (454 Life Sciences (Roche)/FLX; http://www.454.com/). In particular, pyrosequencing has been widely used for SNP detection (Fakhrai-Rad et al. 2002; Langae and Ronaghi 2005; Ronaghi et al. 2007) because it offers an easy, quantitative, and accurate method to detect the relative abundance of alleles in a mixture, provided they differ by a single nucleotide and that their relative frequency is between 10% and 90% (Ronaghi et al. 1998; Wasson et al. 2002).

We realized that with two simple modifications, pyrosequencing could be used to effectively identify changes in copy number at any locus: First, insertion of SNVs into the probes (a reference probe and a query probe) eliminates the reliance on natural variation among alleles. Second, if the reference and query probes are ligated into a single reference/query pyrosequencing (RQPS) probe, their molar ratio will be fixed at 1:1, allowing elucidation of the copy number by pyrosequencing these two SNVs in mixtures of the test DNA and the RQPS probe. We show here that RQPS could easily determine the copy number of any query gene (or DNA fragment) in a genome, be it a man-made transgene, the result of a natural CNV (including both gain and loss), or gene loss as a result of homologous recombination. Unlike previous methods (Pielberg et al. 2003; Soderbäck et al. 2005), RQPS does not require any prior knowledge of the duplication break point or an endogenous reference fragment. It thus offers a flexible and universal CNV determination tool that can be applied to any individual from any species. In addition, it is ideal for the detection of targeted genes in ES cells. The simplicity, automation-friendly protocols, and low cost make it suitable for processing a large number of samples for any specific locus, which could be particularly useful for clinical disease-related CNV screening and patient diagnosis.

In the course of developing RQPS, we established new protocols that enable fast and economic enrichment for homologous recombinant ES cells by simple modifications of the targeting vectors. This is achieved by introducing one SNV into a mouse phosphoglycerate kinase 1 (Pgk1) promoter, which drives the expression of the antibiotic selective cassette, and another into one of the homology arms of the targeted gene. Only male ES cells experiencing homologous recombination would display an equal ratio between the introduced SNVs residing in the targeting construct and the respective endogenous loci for both Pgk1 and the targeted allele (for details, see Fig. 5). A simple pyrosequencing screen (pyrosequencing) could reduce the screening task by removing up to 80% of nonhomologous ES cell clones from the candidate pool. This significantly reduces the effort and increases the number of clones that can be screened effectively for hard-to-target regions. While we demonstrate the utility of RQPS for identification of homologous recombinants in ES cells and show that RQPS, either alone or in combination with pyrosequencing, provides an easy, safe, cheap, BAC and conventional targeting vector compatible ES cell screening method, its real strength is that it offers a powerful and flexible tool for gene copy enumeration in any DNA-containing organism.

Results

Principle of RQPS

Unambiguous identification of positive homologous recombinants (HR), independent of homology arm length, is desired to facilitate the use of BAC targeting vectors. To achieve this, we developed the RQPS method (Fig. 1). The key to this method is the creation of a DNA probe consisting of two fused DNA fragments. The first fragment is designed from a reference gene with known copy number, preferentially one that produces a haploinsufficient phenotype (to ensure both copies are present), and the second from the query gene. Importantly, each of the fragments could be as short as 50 bp and designed to contain a novel SNV to differentiate the probe from the wild-type allele. Such probes can be created by cloning and mutagenesis, by synthesis of long oligos, or by PCR amplification with primers containing desired SNVs. This

Figure 1. Schematic illustration of RQPS. A RQPS probe, which consists of a DNA fragment from the query gene physically linked to a DNA fragment from a reference gene, can be used to detect the relative abundance of alleles in a mixture, provided they differ by a single nucleotide and that their relative frequency is between 10% and 90% (Ronaghi et al. 1998; Wasson et al. 2002). The first fragment is designed from a reference gene with known copy number, preferentially one that produces a haploinsufficient phenotype (to ensure both copies are present), and the second from the query gene. Importantly, each of the fragments could be as short as 50 bp and designed to contain a novel SNV to differentiate the probe from the wild-type allele. Such probes can be created by cloning and mutagenesis, by synthesis of long oligos, or by PCR amplification with primers containing desired SNVs. This

2082 Genome Research www.genome.org
strategy ensures that the two artificial SNVs in the mixture are always at a 1:1 molar ratio.

Figure 1 describes how to determine the copy number of any query gene (endogenous, integrated, or episomal). First, the RQPS probe is added to genomic DNA samples of known concentration. Since pyrosequencing is most accurate in mixtures where the percentage of SNVs is between 10% and 90% (Wasson et al. 2002), two mixtures are generated for each sample. In the first, the molar ratio of the RQPS probe to sample genomic DNA is −1:1, and in the second it is higher (accuracy is not important; more probes may be added when the query copy number is very high). We define the reference SNV as R-SNV and the query SNV as Q-SNV. In this example, assume that the probe R-SNV is a guanine (G), the probe Q-SNV is thymidine (T), the sample reference (R) (genomic or other) is adenine (A), and the sample query (Q) is cytidine (C); then the total (100%) of (R-SNV) + (R) in the mixture (%G + %A) is equal to the total (Q-SNV) + (Q) or (%T + %C). When an autosomal gene with two copies per genome is used, m is defined as the fraction of R-SNV and n is defined as the fraction of Q-SNV in a mixture, the fraction of R (%A) is (1 − %G) or (1 − m) and the fraction of Q (%C) is (1 − %T) or (1 − n). While the sum (R-SNV) + (R) is always equal to (Q-SNV) + (Q), the T/C ratio is equal to the G/A ratio only if the query, like the reference gene, is a diploid locus. T/C would be twice G/A if one query allele is lost (as expected if a homologous recombination has occurred and one copy of the wild-type allele is deleted or replaced with the payload). If many copies of a query transgene (or virus) were integrated, T/C would have a smaller value than G/A. In a more abstract form, the molar ratio of probe to sample in the mixture (G/A above) could be expressed as 2m/(1 − m) for a diploid locus. This ratio is equal to t/tn/(1 − n), where t is the copy number of the query gene in the genome. (If the query is also a diploid locus, t = 2; if more copies of the query are present, then t > 2. See Methods for details.) To derive the value (t), we plotted the G/A values against the T/C values to produce a line with zero intercept and calculate its slope (k). The value (t) is equal to 2k (t = 2k) when the diploid reference locus is used. To increase the accuracy, we plotted at least two m/(1 − m) values obtained in different reaction mixtures against n/(1 − n) values of the corresponding mixture. A graphic representation of these calculations is shown in Figure 1.

Gene copy number determination with RQPS

To demonstrate the utility of this method, we determined the gender of 18 animals in a blind experiment by applying RQPS to assess the copy number of the X-linked gene with two copies per genome is used as a reference, t/tn/(1 − n) is adenine (A), and the sample query (Q) is cytidine (C); then the total (100%) of (R-SNV) + (R) in the mixture (%G + %A) is equal to the total (Q-SNV) + (Q) or (%T + %C). When an autosomal gene with two copies per genome is used, m is defined as the fraction of R-SNV and n is defined as the fraction of Q-SNV in a mixture, the fraction of R (%A) is (1 − %G) or (1 − m) and the fraction of Q (%C) is (1 − %T) or (1 − n). While the sum (R-SNV) + (R) is always equal to (Q-SNV) + (Q), the T/C ratio is equal to the G/A ratio only if the query, like the reference gene, is a diploid locus. T/C would be twice G/A if one query allele is lost (as expected if a homologous recombination has occurred and one copy of the wild-type allele is deleted or replaced with the payload). If many copies of a query transgene (or virus) were integrated, T/C would have a smaller value than G/A. In a more abstract form, the molar ratio of probe to sample in the mixture (G/A above) could be expressed as 2m/(1 − m) for a diploid locus. This ratio is equal to t/tn/(1 − n), where t is the copy number of the query gene in the genome. (If the query is also a diploid locus, t = 2; if more copies of the query are present, then t > 2. See Methods for details.) To derive the value (t), we plotted the G/A values against the T/C values to produce a line with zero intercept and calculate its slope (k). The value (t) is equal to 2k (t = 2k) when the diploid reference locus is used. To increase the accuracy, we plotted at least two m/(1 − m) values obtained in different reaction mixtures against n/(1 − n) values of the corresponding mixture. A graphic representation of these calculations is shown in Figure 1.

Transgene copy number and zygosity determination with RQPS

The frequent use of transgenic Cre lines and the difficulty in determining transgene copy number motivated us to apply RQPS to differentiate homozygous from heterozygous pups, a task that previously required knowledge of the transgene integration site or a test breeding. For this purpose, a general RQPS probe (Notch2-Cre RQPS) was prepared. Pax3-Cre transgenic mice (Li et al. 2000) were chosen for proof-of-principle analysis since the copy number and integration site of this transgene remained unknown. As shown in Figure 3A, we calculated the value (t) for four male pups by plotting the results from different probe dilutions for each genomic DNA sample. We observed three lines with the slope k = −8 and one with the slope k = −16. Multiple rounds of mating these studs with wild-type females confirmed that RQPS identified the lone homozygous male (Fig. 3A). Simultaneously, RQPS estimated the Pax3-Cre transgene copy number in hemizygous to be roughly 16 (Fig. 3A). To determine if the Notch2-Cre RQPS probe could be used as a universal reagent for all Cre transgenic mice, we tested the Tet-O-Cre (Perl et al. 2002) and Sic2-Gfp:Cre (Humphreys et al. 2008) lines and determined that both contained two copies in hemizygotes (Fig. 3B; Supplemental Fig. S1). These results validated RQPS as a rapid, simple solution for the simultaneous determination of transgenic breeder zygosity and copy number. The Notch2-Cre RQPS probe is available upon request.

Detection of copy number variation in human patients using RQPS

We next tested if RQPS could be used to measure the gene copy number in seven human patient samples previously found to have duplication or loss of specific loci with array-CGH. Without knowing if loss of homozygosity (LOH) or duplication had occurred, the samples were subjected, in a semi-blind manner, to RQPS using probes consisting of small DNA fragments from genes located in the aberrant region, as query, and NOTCH2, as reference. The results obtained from five samples by RQPS were consistent with array-CGH data, including all four LOH cases and one duplication (Table 1). However, RQPS results conflicted with array-CGH for two
samples. In both cases, the readout from RQPS was smaller (two copies by RQPS vs. three copies by array-CGH). Since the RQPS probes that we designed did not overlap with array-CGH probes, this could reflect a scenario wherein the region covered by the 60-bp-long array probe is duplicated, whereas the region detected by our RQPS probe is not. Indeed, when RQPS probes were redesigned to match the array-CGH probes, the two methods produced consistent results, identifying duplication in both loci. The discrepancy between the exact copy numbers (three copies vs. four copies) may not be surprising since probe response to ratio change in oligonucleotide arrays tends to be suppressed (Carter 2007). These data demonstrate that RQPS offers an economic alternative for confirmation of array-CGH results and is particularly effective in detecting defined deletions/amplifications. RQPS could be also used to map the boundaries of CNV.

**Screening for homologous recombinants in ES cells**

Next, we used RQPS to determine if we could correctly identify homologous recombination in ES cells by detecting the loss of one allele at the targeted locus. We tried this with ES cells that were electroporated with the Notch1::CreERT2 construct. Homologous recombination will replace the Notch1 intracellular domain with DNA encoding CreERT2 and a FRT flanked selective marker; hence, only one wild-type allele remains in HR clones. We tested 32 ES cell clones enriched for HR candidates using a RQPS probe containing the Notch2 reference fragment and a query DNA fragment corresponding to the intracellular domain of Notch1 (Notch2–Notch1 RQPS) (see enrichment protocol below; Supplemental Table S3). Most clones had k-values of −1 (diploid), but nine clones had k-values near 0.5, as expected if one Notch1 allele is lost (Fig. 4A). A Southern blot of all 32 clones confirmed the identity of the nine clones (Fig. 4B), and RQPS correctly identified all the homologous recombinants in this pool. Therefore, the RQPS screening produced an accurate, simple, BAC-compatible HR identification protocol.

**Enrichment for HR by incorporating SNV(s) into targeting vectors**

RQPS can work as a primary screening method to accurately identify homologous

---

**Figure 2.** RQPS accurately determines gene copy number. (A) Illustration of the Notch2–Pgk1 probe with SNVs introduced to differentiate it from the genomic counterparts. In the RQPS probe, the reference Notch2 fragment has a G-to-C SNV while Pgk1 has a C-to-G SNV. The fragment containing the Notch2 SNV was amplified with a three-primer system: (1) the forward tailed primer F1 with a complementary region to (2) the universal biotinylated primer (UBP), and (3) the untailed reverse primer R1. The fragment containing the Pgk1 SNV was amplified with the same UBP primer and two gene-specific primers (untailed forward primer F2 and a tailed reverse primer R2). The biotinylated PCR product was purified with streptavidin-coated Sepharose beads and sequenced with a sequence primer (S1 for untailed forward primer F2 and a tailed reverse primer R2). The biotinylated PCR product was purified with streptavidin-coated Sepharose beads and sequenced with a sequence primer (S1 for untailed forward primer F2 and a tailed reverse primer R2). The biotinylated PCR product was purified with streptavidin-coated Sepharose beads and sequenced with a sequence primer (S1 for untailed forward primer F2 and a tailed reverse primer R2).
recombinants regardless of homology-arm (HA) lengths, but it is still laborious if one is to screen hundreds of ES cells. To simplify the task and enrich for ES cells that experienced homologous recombination events independent of HA size, we introduced another simple innovation: insertion of two SNVs to the targeting vector (Fig. 5A). As mentioned in the introduction, one SNV is placed on the homology arm (the HA-SNV), roughly 50 bp away from the payload; the other was inserted into the neomycin selection cassette, using the same \( \text{Prgk} \) promoter SNV described in Figure 2. Enrichment for putative HR is accomplished by simply examining the ratio of these SNVs to their endogenous counterparts in a single quantitative pyrosequencing reaction (Fig. 5A). Because most facilities use male (XY) ES cells, only ES cell clones that have experienced a single integration, homologous or random, would display an allele ratio of 1:1 for the X-linked \( \text{Prgk} \) SNV. Any additional integration will increase the proportion of the vector \( \text{Prgk} \) SNV (two integration 66%, three 75%, etc.). On the other hand, the HA-SNV will only show a 1:1 ratio in homologous recombinants or in cells that experienced two random integrations, while under all other conditions, the fraction of HA-SNV will fall either around 33% (one random integration) or above 60% (at least three integrations). Therefore, only HR will display a 1:1 ratio for both SNVs. We called this enrichment method “pyroscreening.”

As a proof-of-principle, ES cells were electroporated with one of five different gene targeting constructs (a detailed description of each line will be provided elsewhere). Four of the five vectors were constructed in \( \text{CS7/B6} \) BAC clones using the BAC recombineering method (Warming et al. 2005), during which the desired SNVs were incorporated (Supplemental Fig. S2). Although pyrosequencing

### Table 1. CNV analysis in human patient samples by RQPS

<table>
<thead>
<tr>
<th>Patient</th>
<th>Genes</th>
<th>RQPS probe (query)</th>
<th>Estimated copy number by RQPS</th>
<th>Estimated copy number by Array-CGH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MAP3K12; SP1; KRT18; ZNF740</td>
<td>SP1 exon3</td>
<td>2(^b)</td>
<td>3</td>
</tr>
<tr>
<td>1(^a)</td>
<td>MAP3K12; SP1; KRT18; ZNF740</td>
<td>SP1 CGH probe region</td>
<td>4(^a)</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>TRABD; MAPK12; FAM19AS</td>
<td>TRABD</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>SMYD1; FABP1; RPIA; EIF2AK3</td>
<td>FABP1 exon1</td>
<td>2(^b)</td>
<td>3</td>
</tr>
<tr>
<td>3(^a)</td>
<td>SMYD1; FABP1; RPIA; EIF2AK3</td>
<td>FABP1 CGH probe region</td>
<td>4(^a)</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>RX3; GLIS3</td>
<td>RX3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>RX3; GLIS3; JAK2; PSIP1; MTAP; BRCA2; SMAD9; SACS; TRN5F19</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>TP53; TNK1; NLGN2; ATP1B2; FGFI1; SOX15</td>
<td>FGFI1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>RAB12; ARHGAP28; RALBP1; NAPG</td>
<td>RALBP1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

For each sample, an RQPS probe, consisting of an exon fragment (~50 bp) from within a randomly selected query gene residing within the CNV region fused to a \( \text{NOTCH2} \) fragment, was prepared. Both fragments were mutated to contain an SNV. These probes were then used to estimate the copy number of seven loci in a semi-blind test. After RQPS, the results were compared to previously determined array-CGH data. Two discrepancies were found, which were resolved when RQPS was repeated with query sequences matching the CGH probe.

\(^a\)Repeat reaction with RQPS probe matching the CGH probe.

\(^b\)Inconsistent result between RQPS probe to a region not covered by a CGH probe and the CGH array.

\(^c\)Both RQPS and array-CGH data detected amplification; they differ in estimating copy number.

---

**Figure 3.** RQPS differentiates homozygous from heterozygous transgenic animals. (A) RQPS on four Pax3-Cre animals. In this experiment, four different dilutions of RQPS probe were made with genomic DNA samples from each mouse; this was done to demonstrate that any two concentrations would have worked as well. Note the very small standard deviations for each data point, indicating that triplicates or duplicates are not necessary. Pyrosequencing of each mixture was performed in triplicate, and the average values were plotted; the standard deviation of each data point is shown. (B) RQPS of Tet-O-Cre mice. The copy number of the Tet-O-Cre transgene was measured in four animals. Four different mixtures were pyrosequenced in duplicate for each sample. RQPS detected two homozygotes and two heterozygotes, with about two and four copies of the transgene, respectively.
for HR was attained by selecting clones with an ~50% SNV ratio for both Pgk1 and HA (Fig. 5B,C). Comparison with traditional PCR and Southern blot screening of the entire population determined that only two of the 21 positive HRs were missed by pyrosequencing (Fig. 5B; Supplemental Figs. S3 and S4). Pyrosequencing is thus an effective enrichment protocol, streamlining the analysis process by eliminating ~80% of all G418-resistant clones prior to RQPS, PCR, or Southern blot analysis with an estimated false-negative rate of ~9.5% (Supplemental Table S3). The false-positive clones are most likely due to the contamination of MEF cells, which affect the ratios of both HA and Pgk1 SNVs. We expect fewer or no false-positives if ES cells are grown without MEF cells. The SNV-containing selection cassette we have created (eukaryotic Pgk1/prokaryotic EM7 dual promoter-driven neomycin/kanamycin flanked by FRT site) is available upon request and should facilitate pyroscreening.

Discussion

We report here the adaptation of quantitative pyrosequencing to measure the copy number for any gene or DNA fragment in any species and its use in detecting homologous recombinants, transgene copy number, zygosity, and CNV validation. In order to apply pyrosequencing to any sequence, the investigator needs to generate a RQPS probe consisting of two SNV-containing fragments with homology with a query locus and a reference locus, respectively. Query loci could be integrated DNA or endogenous loci that underwent micro duplication or loss of DNA. The design of the SNV-containing RQPS probes renders our method independent of naturally occurring SNPs and reference alleles, distinguishing it from other pyrosequencing-based CNV detection methods reported so far (Pielberg et al. 2003; Söderhäll et al. 2005). This endows RQPS with great flexibility while retaining the simplicity and cost effectiveness of pyrosequencing. Once a probe is made for a specific locus of interest, a large number of samples could be easily and economically processed, which makes RQPS ideal for high-throughput clinical diagnosis of the CNV status at a limited number of loci.

Another application we demonstrated is an enrichment protocol to identify potential homologous recombinants. By introducing two artificial SNVs, one into the Pgk1 promoter of the selection cassette and the second into the homology arm, most negative clones can be discarded (Fig. 5). It is important to point out that the MEF feeder cells used in this study are derived from the F1 embryos of a cross between Jackson transgenic C57BL/6-Tg (pWLS12hgy1Ems/J) homozygotes and C57BL/6-Tg (pPGKnheo/bpa3Ems/J) homozygotes. In such cells, both hygromycin- and neomycin-selective cassettes are driven by mouse Pgk1 promoters. Assuming an equal number of male and female MEF cells, at least one copy from each transgene and 1.5 copies from the endogenous Pgk1 gene are present, and thus each MEF cell contains at least 3.5 copies of Pgk1 promoter. Despite this confounding factor, we achieved an overall threefold enrichment even with a single SNV pyroscreening. Better performance of pyroscreening will be attained when ES cells are grown without feeder cells (Nichols et al. 1990). It is worth noting that placing a homology-arm, silent SNV into an exon in the gene-targeting construct offers the important additional benefit of distinguishing the two transcripts generated from the endogenous and targeted alleles by quantitative pyrosequencing on reverse-transcribed cDNA.

In contrast to pyrosequencing, ES cells can be directly screened with RQPS independent of targeting vectors. Like Q-PCR, RQPS screens for positive clones by detecting the loss of one wild-type allele in the targeted locus (Valenzuela et al. 2003). Yet unlike Q-PCR, RQPS does not measure a twofold reduction in the amount, but rather determines the ratio of two different nucleotides. Because the relative ratio of alleles in a mixture is measured within a single tube, system errors are minimized. We have shown that RQPS can correctly identify all positive clones after enrichment by pyrosequencing and could thus replace the Southern blot screening method. As BAC recombinering has made the construction of a large number of gene targeting vectors in a short period of time possible (Valenzuela et al. 2003), the automation-friendly RQPS, alone or in conjunction with pyrosequencing, will effectively streamline the ES cell screening process.

Methods

Generation of targeting constructs

The five targeting constructs described in the text are: (1) Notch1::Cre<sup><sup>ER</sup></sup>; (2) Notch1::CreERT2, in which the Notch1 intracellular domain (ICD) is replaced with Cre recombinase or tamoxifen-inducible CreERT2 fusion protein, respectively; (3) Notch2::cre, in which the Notch2 ICD is replaced with Cre recombinase; (4) N2-N1, in which the Notch2 ICD is replaced by the Notch1 ICD; and (5) N1-N2 in which the Notch1 ICD is replaced by the Notch2 ICD. To explain how SNVs are simultaneously incorporated into a targeting construct through BAC recombinering, we provide details of constructing the N1::CreERT2 vector (below; Supplemental Figure 4).
Electroporation of ES cells with (black dots) constructs in ES cells (HR, homologous recombination; RI, random integration) yield different ratios of mouse SNVs in front of a FRT site (Supplemental Fig. S2). This strategy (insertion of the mini-HR) circumvents the need to amplify the payload in order to introduce homology arms into it and thus facilitates the insertion of large DNA fragments into a BAC. For the studies described here, the CreERT2FRTNeo cassette with 4.8 kb upstream and 2.2 kb downstream flanking BAC DNA was recombined into a pBluescript plasmid as described (Liu et al. 2003).

ES cell electroporation

Forty micrograms of linearized targeting construct was electroporated into either 129 derived SSC10 (for N1::creFRT) or C57BL6 derived B6/BLU (all other constructs) ES cells by the Siteman Cancer Center ES Core at Washington University (http://escore.im.wustl.edu). Neomycin-resistant clones were selected by the core and expanded by the investigators for further screening.

PCR and Southern blot screening of ES cells

Nested PCR was used for the screening of ES cells that were electroporated with targeting vectors with long HAs. Purified BAC DNA containing a desired targeting construct was diluted with wild-type mouse genomic DNA and used as a positive control for the optimization of PCR conditions. PCR primers with the expected size of the PCR products are listed in Supplemental Table S4. The 25-μL PCR reaction included: 2.5 μL of 10× PCR buffer, 0.25 μL of KlenTaqLA (Sigma), 0.5 μL of 10 mM dNTP mixture, 1 μL each of 20 μM upstream and downstream primers, 8.25 μL of 4 M betaine, and 1 μL of DNA template of unknown concentration. The PCR cycle conditions were 94°C (2–5 min depending on the amplicon size at 1 min/kb), 35 cycles of 94°C (1 min), 50°C–65°C (1 min), and 68°C (2–5 min).

Pyrosequencing

For each measurement, we PCR-amplified the SNV-containing region with two target-specific primers flanking the desired SNV, one of which (a tailed primer) included a universal sequence complementary to a third, the universal biotinylated primer (UBP) (Fig. 2A; Fakhrai-Rad et al. 2002; Royo et al. 2007). All primers were designed using Pyrosequencing Assay Design Software (Biotage) (Supplemental Table S2). It is worth noting that the PCR amplification efficiency of fragments that differ by only a single SNV in the same PCR reaction mixture is the same unless the SNV creates or disrupts a stable secondary structure, a complication that can be avoided when the SNV is chosen by the investigator. Single-strand

Rapid determination of gene copy number

In brief, a 6xMyc-tagged CreERT2 fusion gene was first cloned in front of a FRT-flanked neomycin selection cassette, whose expression was under the control of both mouse Pgk1 and Escherichia coli EM7 promoter. An artificial SNV was introduced into the mouse Pgk1 for pyrosequencing by site-directed mutagenesis. Next, we amplified a fragment containing the blastidin promoter-driven ampicillin-resistance gene using pBlueScript as a PCR template. Two pairs of flanking mini-HA were added in subsequent rounds of PCR. The outer mini-HAs (magenta) mediated the recombination of the fragment into a Notch1 BAC; the inner mini-HAs (cyan) mediated subsequent replacement of the ampicillin cassette by recombination of a payload cassette (6xMyc-CreERT2FRT; PGKEM7-Neo). In addition, a second SNV was inserted into the outer mini-HAs (Supplemental Fig. S2). This strategy (insertion of the mini-HR) facilitates subsequent replacement of the ampicillin cassette by recombination of a payload cassette (6xMyc-CreERT2FRT; PGKEM7-Neo). In addition, a second SNV was inserted into the outer mini-HAs (Supplemental Fig. S2). This strategy (insertion of the mini-HR) facilitates subsequent replacement of the ampicillin cassette by recombination of a payload cassette (6xMyc-CreERT2FRT; PGKEM7-Neo).
DNA PCR products are then purified with streptavidin-conjugated Sepharose beads and used for pyrosequencing analysis. The whole procedure is high-throughput-friendly, and 96 samples could be processed within 3–4 h (from the setup of PCR reactions to the obtaining of pyrosequencing results). This three-primer PCR reaction system (Fakhrai-Rad et al. 2002; Royo et al. 2007) was used for the amplification of the DNA fragments containing an SNV in a PCR reaction mixture containing 2 μL of DNA sample, 12.5 μL of Bullseye Taq DNA Polymerase Mix (Midsci), 1 μL of untailed gene-specific primer, 0.2 μL of tailed gene-specific primer, 0.8 μL of UBP-labeled (all at a concentration of 5 μM), and 8 μL of distilled water (primer sequences are provided in Supplemental Table S2). PCR conditions were 8 min at 95°C followed by 40 cycles at 95°C (30 sec), 60°C (30 sec), and 72°C (30 sec).

Single-strand PCR products were purified as described (Bruinsma et al. 2008). Briefly, single-stranded biotinylated PCR products were immobilized on streptavidin-coated Sepharose beads in a vacuum prep tool (Biotage). For each reaction, 40 μL of Binding Buffer (10 mM Tris-HCl, 2 M NaCl, 1 mM EDTA, 0.1% Tween 20 at pH 7.6), 20 μL of double-distilled water, and 3 μL of Streptavidin Sepharose High Performance beads (GE Healthcare) were mixed together with 20 μL of PCR product for 10 min at room temperature using an Eppendorf Thermomixer. After immobilization, vacuum was applied to the prep tool and the template was captured on filter probes. The non-biotinylated template was then denatured by alkali treatment and removed by washing for 5 sec in 70% EtOH, 0.2 M NaOH denaturation solution (5 sec), and 10 mM tris-acetate washing buffer (pH 7.6) for 5 sec. The last washing step neutralized the pH. Next, the vacuum was released, and the beads were transferred into a Biotage PSQ 96 well plate containing the Sequencing Primer (5 pmol) and 40 μL of annealing buffer (20 mM tris-acetate, 2 mM MgAc2 at pH 7.6). The pyrosequencing reactions were performed using a PSQ96-MA machine (Biotage) according to the manufacturer’s instructions using the Pyro Gold reagent kit (Biotage). Assays were performed using cyclic enzyme and nucleotide dispensation, followed by analysis using the SNV Software (Biotage).

ES cells from individual clones were trypsinized and harvested. Genomic DNA was purified as described (Laird et al. 1991) and dissolved in 50 μL of TE (pH 8.0). For pyrosequencing, 2 μL of ES cell genomic DNA (unknown concentration) was used for PCR reaction, followed by single-strand DNA purification and pyrosequencing. Clones with an SNV ratio between 40% and 60% were considered as positive candidates for further screening.

**RQPS**

For RQPS, the concentration of both genomic DNA and the line-arized RQPS probe was measured. At least two mixtures of genomic DNA/RQPS probe were then prepared, and the molar ratio between the two was between 9:1 and 1:9 (we used 1:1 and 1:2 when two mixtures were done). When 30 ng of mouse or human genomic DNA is used, for an RQPS probe with the size of 3–4 kb, 0.03–0.04 pg is required to get a roughly 1:1 molar ratio. For the detailed calculation method, please see http://www.med.umich.edu/tamc/spike.html. A 2-μL sample from each mixture was used as the template for the PCR reactions with reference or query specific primers, followed by single-strand DNA purification and pyrosequencing.

**Mathematical derivation of gene copy numbers with RQPS**

To calculate the copy number of a query gene in a genome (f), we created a mixture containing (a) genomes and (b) molecules of RQPS probe. Since each RQPS probe contains one reference sequence, there are (b) R-SNV nucleotides in the mixture. In contrast, each diploid genome contains two alleles of the autosomal reference gene and (f) copies of the query gene. Thus, the number of Rs-SNV nucleotide is (2a) and the number of Qs-SNV is (fa). Therefore, the fraction of the R-SNV in the mixture (m) is equal to the molar ratio R-SNV/(R-SNV + Rs-SNV) or:

\[ m = \frac{b}{2a + b}. \]  (1)

Likewise, the fraction of the Q-SNV in the mixture (n) is

\[ n = \frac{fa}{a + fa}. \]  (2)

Equation 1 could be solved for \( b \):

\[ b = 2 \times a \times m/(1 - m) \]  (3)

and, similarly, Equation 2 could be solved for \( b \):

\[ b = t \times a \times n/(1 - n). \]  (4)

Thus,

\[ 2 \times a \times m/(1 - m) = t \times a \times n/(1 - n) \]  (5)

and

\[ t = 2m(1 - n)/(1 - m)n. \]  (6)

When the value \( m/(1 - m) \) is plotted against \( n/(1 - n) \), a line passing the origin will have the slope \( k \) that equals half of the transgene copy number in the genome \( k = 0.5t \). If a single-copy gene is used as reference, \( k = t \).

For the detection of the loss of one allele in ES cells, contamination from feeder cells needs to be considered. If MEF cells are present in cases in which only nonhomologous recombination occurs, the copy number of targeted alleles per genome will remain 2. If the ES cell undergoes homologous recombination, there is only one targeted allele in each genome of the ES cell, whereas two copies of the targeted allele are present in the MEF genome. This will result in higher \( t \) values than the expected 0.5, resulting in a false-negative.

**Data process**

All pyrosequencing data were processed with Microsoft Excel. \( m/(1 - m) \) was plotted against \( n/(1 - n) \), and linear data trend lines were forced to have a zero intercept.

**Generation of animals with defined copy number of Notch1 exon 30**

To generate animals with one, two, three, and four copies of exon 30 but with a diploid number of the Notch1 extracelluar domain (exon 23), we used the N1::Cre line (Vooijs et al. 2007) in which Cre replaced exon 30 in one allele, and the Gt(Rosa)26Sor^Notch (Murtaugh et al. 2003) line in which one copy of the Gt(Rosa)26Sor locus is replaced with a cDNA encoding N1ICD to generate a mouse with three copies of exon 30. Wild-type animals have two alleles of Notch1, and homozygous N1^+/–; Gt(Rosa)26Sor^Notch/Notch has four copies of exon 30.

**Array-CGH**

Array-CGH was carried out as described (Schwarzbraun et al. 2009).

---

Liu et al.