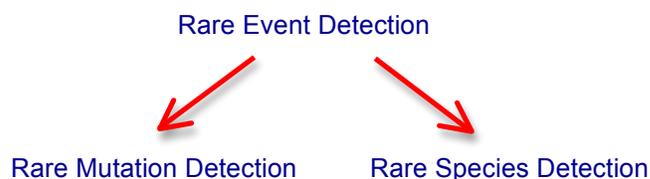




Rare Event Detection

October 1, 2012

Rare event detection (RED) encompasses any type of application that is focused on the lower limits of detection, particularly in the context of high amounts of background DNA. RED can be separated into two classes: Rare Mutation Detection (RMD) and Rare Species Detection (RSD).



	Measurement Units	Examples	Assay Components	Assay Schematic
Rare Mutation Detection (RMD)				
	Ratio (% or a/(a+b))	<ul style="list-style-type: none"> Tumor biopsy: 0.01% mutant detected, or 1 in 10,000 SNPs Small indels 	1 primer pair, 2 competitive probes	
Rare Species Detection (RSD)				
Case 1	Copies/volume	<ul style="list-style-type: none"> Invasive species monitoring (copies/mL lake water) HIV detection (copies/mL plasma) 	1 assay (1 primer/probe set)	
Case 2	Ratio (% or a/(a+b)) or Copies/unit reference	<ul style="list-style-type: none"> Gene expression of rare transcripts Viral staging indels 	2 independent assays (2 primer/probe sets, run either as singleplex or duplex)	

Figure 1: Rare Event Detection includes both Rare Mutation Detection (RMD) and Rare Species Detection (RSD). RMD and RSD can be classified according to their assay components. An RMD assay is comprised of a single primer set plus two competitive probes. RSD utilizes either a single assay to detect a rare species per unit volume (Case 1) or two independent, non-competitive assays to detect a rare species as a proportion of a more common species (Case 2). Some applications may combine aspects of RMD and RSD.



Rare Mutation Detection

Challenge

An RMD assay must discriminate between two highly similar sequences, one of which is significantly more abundant than the other. An example of an RMD assay is detection of a single nucleotide polymorphism (SNP) in a cancer biopsy.

Assay Components and Considerations for RMD

RMD can be defined and distinguished from other types of rare event detection based on its assay components. In ddPCR an RMD assay is typically composed of a single set of primers plus two competitive probes, one detecting the wild-type allele, and one detecting the variant allele (Figure 1). In the case of SNP allelic discrimination assays, these two probes may differ by only a single nucleotide.

In a bulk reaction, such as standard qPCR, probe cross-reactivity is the greatest hindrance to sensitive allelic discrimination, with sensitivity typically limited to around 1 in 100, or 1%. By partitioning the DNA sample into droplets, ddPCR effectively increases the mutant to wild-type ratio in the subset of droplets that contain a mutant. This reduces the impact of probe cross-reactivity and provides higher sensitivity and accuracy.

Statistical Considerations for RMD Experimental Design

Limit of Detection

The limit of detection (LoD) is the minimum concentration of the mutant (rare) species that can be reliably differentiated from a negative control (100% wild type). In RMD, LoD is typically quoted as a ratio or a percentage: for example, 1 in 10,000 or 0.01%. In qPCR the LoD is largely a function of the cross-reactivity of the probes. In contrast, in ddPCR the LoD is determined primarily by the number of wild-type molecules that are screened. The LoD can simply be adjusted to the precise requirement of any application by adjusting the number of wells, and therefore molecules, screened.

Determining the Number of Molecules to Screen: The 3X Rule

Consider the case of a sample at a mutant concentration of 1 in 10,000. In order to guarantee with 95% confidence that at least 1 mutant molecule will be detected in this sample, statistics dictates that at least 30,000 wild-type molecules need to be screened, or 3 times the number of expected wild-type target molecules.

$$\# \text{ of Molecules to Screen} = 3 \times \text{Background Molecules (i.e. wild type)}$$

Keep in mind that additional wells may be screened to ensure detection of more than one positive droplet. The number of cells or amount of DNA needed to screen a given number of background molecules is described in Table 1.

Experimental Strategies for RMD

Once the number of target molecules to screen has been determined for a given LoD, the number of wells required to screen that number of targets must be determined. In the QX100 ddPCR system, up to 75,000 molecules may be screened in one well, for a LoD of 1 in 25,000. To achieve a lower LoD more wild-type molecules can be screened by running the same sample in multiple wells, as illustrated in Table 1.



Material to screen (human)				
LoD	Total copies to screen	Diploid cells	Amount of DNA to screen	# of wells
1 in 1,000	3,000	1,500	0.010 ug	1
1 in 10,000	30,000	15,000	0.10 ug	1
1 in 25,000	75,000	37,500	0.25 ug	1
1 in 100,000	300,000	150,000	1.0 ug	4
1 in 1,000,000	3,000,000	1,500,000	10 ug	40

Table 1 Requirements for different limits of detection (LoDs) using the QX100 ddPCR system, assuming an ideal assay with a droplet false positive rate of 0. In order to screen the specified amount of DNA, slightly more DNA must be loaded into a well to account for sample loss. The row in bold reflects the maximal performance in one well. Researchers may choose to screen additional wells to ensure detection of more than one positive droplet.

Real World Factors that Impact RMD Calculations

The table above outlines RMD experimental design based on statistical requirements. Additional real-world factors will impact these numbers in an assay-specific and sample-specific way and should be taken into account when planning an RMD experiment.

- ∞ **Sample Availability:** Most researchers are limited by the amount of clinical sample at their disposal. This is a critical consideration in establishing the LoD. For example, with a typical clinical sample of 100 ng of DNA, the best achievable limit detection is 1 in 10,000, regardless of platform.
- ∞ **Sample Integrity and Preparation:** The calculations above assume that every target present can be successfully assayed. In reality, samples have varying degrees of degradation, and certain sample matrixes can introduce PCR inhibitors. FFPE samples, for instance, are routinely used in cancer research, but provide DNA of poor quality. Sample quality and preparation will impact the amount of starting material needed to effectively screen the correct number of target molecules.
- ∞ **False Positive Rate:** Most systems have some droplet false positive rate due to a combination of instrument, signal processing, and assay effects. While the false positive rate can be very small for a well-designed assay, it is important to take it into account in defining the requirements for a given detection threshold. The false positive rate of an assay can be estimated by running the assay on a 100% wild-type sample.



Rare Species Detection

Challenge

An RSD assay must detect extremely low levels of a given target DNA in a large pool of background DNA. Accurate quantification of the target at very low concentrations is often required. The target species can be any sequence that is markedly different from the background or reference species.

Assay Components and Considerations for RSD

An RSD application might require one or two assays. If the goal is to quantify the amount of target species in a given volume of starting material, a single assay is used (Figure 1: RSD Case 1). If the goal is to quantify the amount of target species with respect to a given reference species, two independent assays are required (Figure 1: RSD Case 2). In either case, assay design and optimization for RSD is in general easier than for RMD because there is less risk of cross-reactivity of the target assay with background DNA.

Statistical Considerations for RSD Experimental Design

Some RSD applications will require reliable quantification of rare species, while others simply require detection of a rare species. This difference will dictate the lower bounds of sensitivity for a given assay and application.

Limit of Detection

For RSD, the LoD can be defined either in terms of the total volume of material analyzed or in terms of the number of copies of some type of background DNA.

Absolute Quantification and Limit of Quantification

In RSD, absolute quantification of the target species is often required. In general, ddPCR is able to provide an accurate quantification of a rare target species at a much lower concentration than qPCR. This is because unlike qPCR, a ddPCR measurement is a count of individual target molecules. This eliminates the need for absolute standards and standard curves and improves reproducibility across experiments and laboratories. In addition, in ddPCR, small variations in PCR efficiency between wells have no effect on the measured concentration. Day-to-day, instrument-to-instrument and researcher-to-researcher measurements are highly reproducible.

When absolute quantification is applied to an RSD application, an important consideration is the limit of quantification, or LoQ. The LoQ is defined as the lowest concentration you can reliably measure within a pre-determined variance, or CV. For example, if you want to quantify within 20%, the LoQ is the lowest concentration that you can reliably measure to within +/- 20% of the real value.

Determining the Number of Molecules to Screen: The 3X Rule

For RSD, the 3X Rule described above for RMD still applies. For example, to reliably detect 1 in 100,000 cells, at least 300,000 background cells must be screened. Similarly, to achieve a LoD of 1 in 1mL of sample, 3mL of sample must be screened.



Experimental Strategies for RSD

Two different experimental setups are recommended for RSD applications depending on whether detection is with respect to a starting volume of sample (Figure 1 Case 1) or with respect to a second background target that requires quantification (Figure 1 Case 2). These approaches are illustrated in the following examples.

Case 1: Quantification with respect to a total starting volume

Example: Detect 1 copy of foreign DNA in 5 mL of lake water

To reliably detect one foreign species of DNA per 5 mL of lake water, at least 15 mL of lake water need to be screened. A schematic of the experiment is shown in Figure 2 for the case of a single ddPCR well. Depending on how much total DNA is extracted from the 15 mL sample, more than 1 well may be needed to analyze the sample. Up to 1.6 μ g of digested DNA can be loaded into one reaction well.

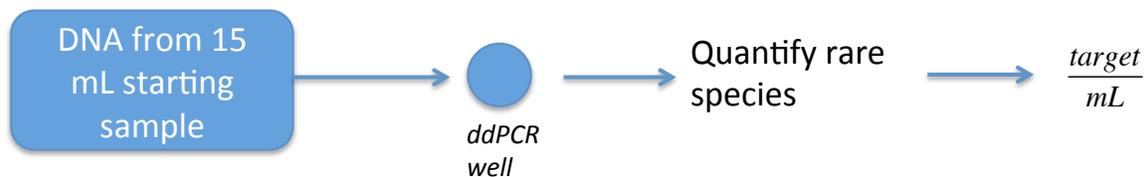


Figure 2: Strategy for detecting rare foreign DNA in a sample of defined volume

Case 2: Quantification with respect to a second DNA species

Example: Detect 1 copy of virus per 100,000 peripheral blood mononuclear cells (PBMCs)

To detect a very low target concentration in a high background sample, the sample can be analyzed at two different concentrations (Figure 3). The high concentration wells provide sensitive detection of the rare species while the low concentration well enables quantification of the background species. For example, to reliably detect an average of 1 virus in 100,000 PBMCs, we need to screen at least 300,000 PBMCs, or 2 μ g of DNA (Table 2). This requires 3 wells on the QX100: two wells at high concentration to screen a sufficient number of PBMCs for the presence of rare virus, and a third well at a lower concentration to quantify the PBMCs. This experimental setup is shown schematically in Figure 3.

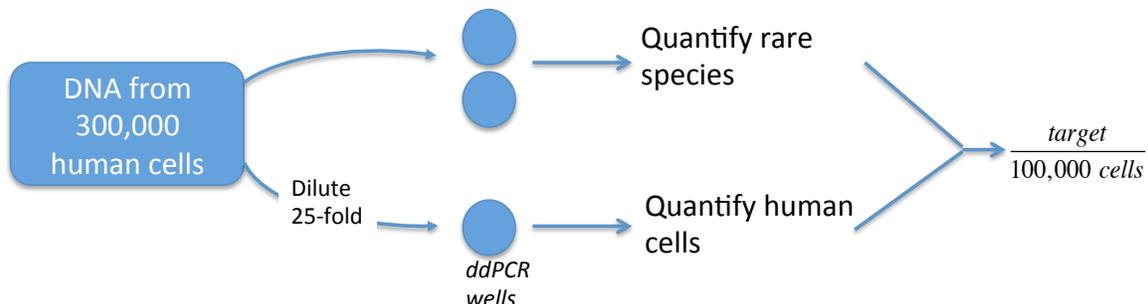


Figure 3: Detection of a very low target concentration in a high background (such as a low viral load in a given number of PBMCs) by analyzing samples at two different concentrations. The target species (the “rare”) species is quantified using a singleplex assay in the high concentration well. The background species is quantified in the low concentration well.



LoD	LoQ (<10% CV)	Human diploid cells	Amount of DNA	# of wells	
				High conc.	Low conc.
1 in 5,000 cells	100 in 15,000 cells	15,000	0.10 ug	1	0
1 in 12,500	100 in 37,500 cells	37,500	0.25 ug	1	0
1 in 100,000 cells	100 in 300,000 cells	300,000	2.0 ug	2	1
1 in 500,000	100 in 1,500,000 cells	1,500,000	10 ug	10	1

Table 2: Experimental setup for different limits of detection (LoD) and limits of quantification (LoQ) in rare species detection, case 2. In the examples shown here, limit of detection is defined with respect to the number of human cells. In order to screen the specified amount of DNA, slightly more DNA must be loaded into a well to account for sample loss.

Example: Accurately quantify (+/- 10%) the amount of virus in 300,000 PBMCs.

In the example given, what would be the limit of quantification? In this situation, the measurement error (CV) is introduced at the time of sub-sampling (at the time of the blood draw), not by the measurement system. Poisson statistics dictate that if N target molecules are measured in a sample, the standard deviation on the measurement is roughly the square root of N. In this example, if 100 copies of virus are present in the sample, the standard deviation of the measurement is 10 and the CV is 10/100 = 10%. This means that at least 100 copies of virus must be detected to achieve an LoQ of +/- 10% CV. No detection system can do better than this theoretical limit.

Real World Factors that Impact RSD Calculations

- ∞ **False Positive Rate:** As with RMD, the false positive rate must be measured and taken into account when designing experiments to achieve very low limits of detection or accurate quantification.
- ∞ **Sample Integrity and Preparation:** Same as with RMD. In the experimental design described in Figure 2, it may be useful to spike a PCR control into the original sample and use a second assay in the same well to measure its level. This enables you to control for PCR inhibitor and sample prep failures.
- ∞ **Pipetting accuracy in dilutions:** If a 100-fold dilution is used to quantify background, as described in Figure 3, the errors introduced by pipetting will contribute to the overall accuracy of the measurement. When used properly for a 100-fold dilution, well-calibrated Rainin pipettes may introduce an uncertainty of roughly 2.5%.