Pyromaker Tutorial

Introduction

DNA, the carrier of genetic information in all living organisms, is the molecule of life. A molecule of DNA is a double helix of two single strands of DNA oriented in opposite directions, where the two single strands anneal because of complementary base pairing between nucleotide bases adenine and thymidine and bases guanine and cytosine. Each single strand is a polymer of individual nucleotides. DNA sequencing technology determines the order of the nucleotide bases in a molecule of DNA and has been used clinically to identify many human genomic disorders that result from DNA mutations. Cystic fibrosis, a chronic lung disease commonly caused by the deletion of three nucleotides from the CFTR gene, is an example of a human genomic disorder that was identified by DNA sequencing.

The online program Pyromaker functions as a virtual model of pyrosequencing, a sequence-by-synthesis method of DNA sequencing. It is distinct from classic Sanger sequencing. Users input wild type and mutant DNA sequences and an order that the user wishes nucleotides to be dispensed, and the software generates the expected resultant pyrogram. In this tutorial, we cover the basis of DNA synthesis and show how Pyromaker demonstrates the fundamental principles of pyrosequencing, specifically how:

1) a peak is generated only when the correct nucleotide is incorporated.

2) peak heights are proportional to the number of nucleotides incorporated.

3) different dispensation sequences can produce different pyrograms.

4) how non-optimized dispensation sequences can produce complex pyrograms.

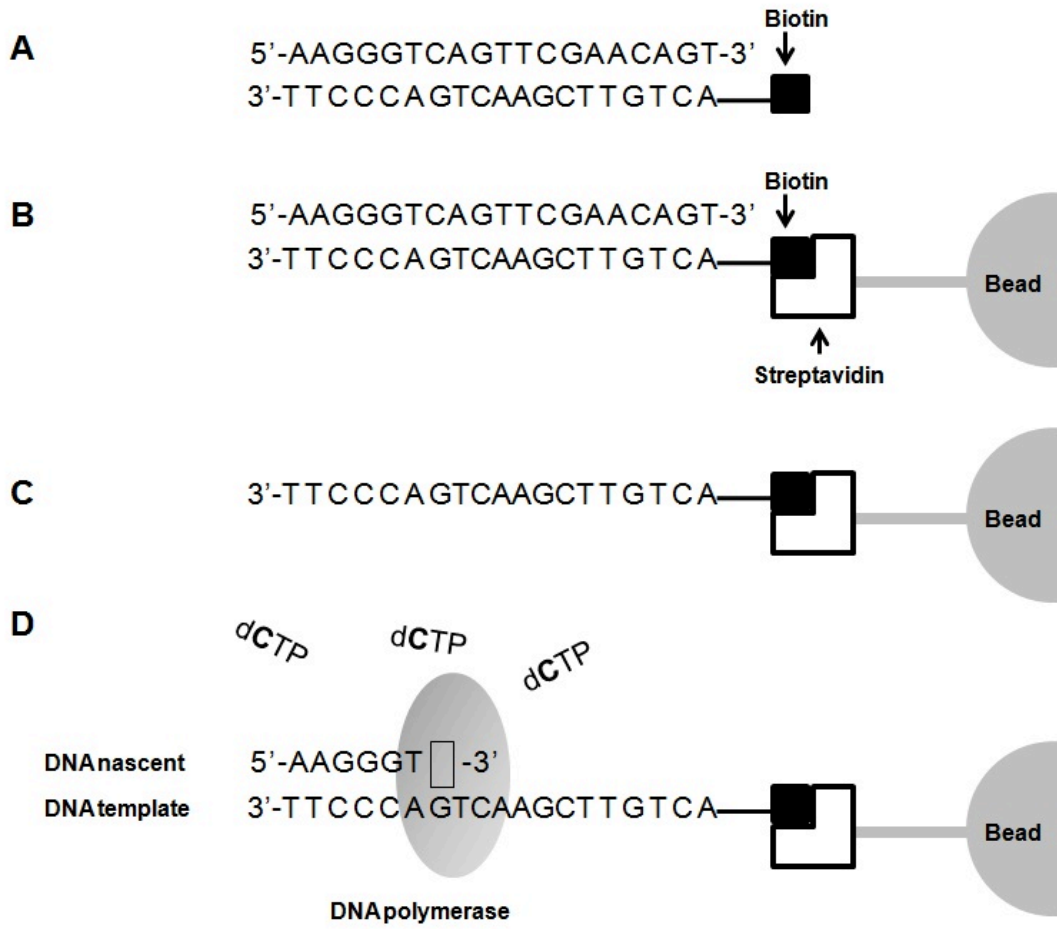
DNA Synthesis

DNA polymerase is the enzyme that replicates DNA in living cells and is also essential for DNA sequencing. Each strand of DNA has a designated 5' and 3' end. The two DNA strands of the DNA double helix are antiparallel, meaning that one strand moves 5' to 3' left to right while the other strand moves 3' to 5' also left to right. DNA polymerase can only extend an already existing strand of DNA when it is bound to a template strand. Thus, primers, which are short DNA fragments, must first pair with the template strand in order for DNA synthesis to occur. On the free end of the primer on the nascent strand is a 3'-OH group. The nascent strand is defined as the growing strand to which nucleotides added to the 3' end (also referred to as the 5' to 3' direction), where the 3' OH is the functional group that indicates where DNA polymerase should begin adding nucleotides. During DNA synthesis, nucleotides are added as dNTPs. The T nucleotide is added to the nascent strand when the polymerase encounters an A on the template strand, and a G is added when a C is encountered on the template strand, and so on.

Pyrosequencing requires four enzymes. Prior to pyrosequencing, the DNA sample is amplified via PCR and the primer for the antisense strand of the PCR products is biotinylated to bind to streptavidin (Figure 1a). For example, streptavidin coated superparamagnetic beads bind to biotinylated DNA through non-covalent bonding (Figure 1b). This non-covalent bond is represented through a lock and key model similar to the binding mechanisms of substrates and enyzmes, yet has a higher affinity than antibody bonds. Once the binding of biotin to streptavidin has occurred, the DNA is denatured, leaving only the single bottom DNA strand immobilized on the beads while the other strand is washed away (Figure 1c). The strand attached to the bead becomes the template strand and undergoes the four enzyme pyrosequencing reaction (Figure

1d). Thus, the complementary nascent DNA strand is synthesized from the template strand attached to the bead. An advantage of using the antisense strand as the template strand is that the DNA sequence produced is sense sequence.

Figure 1



DNA polymerase catalyzes the synthesis of DNA by extending the 3' end of the complementary strand. When the correct complementary dNTP is added, pyrophosphate is released and the reaction, catalyzed by DNA polymerase, is as follows, where n is the number of nucleotides in the nascent strand.

$$(DNA)_n + dNTP \rightarrow (DNA)_{n+1} + PP_i + H^+ \quad \text{(equation 1)}$$

Through a sequence of another two reactions, pyrophosphate is converted into ATP, which is then a substrate for the enzyme luciferase, oxidizing luciferin to oxyluciferin and light.

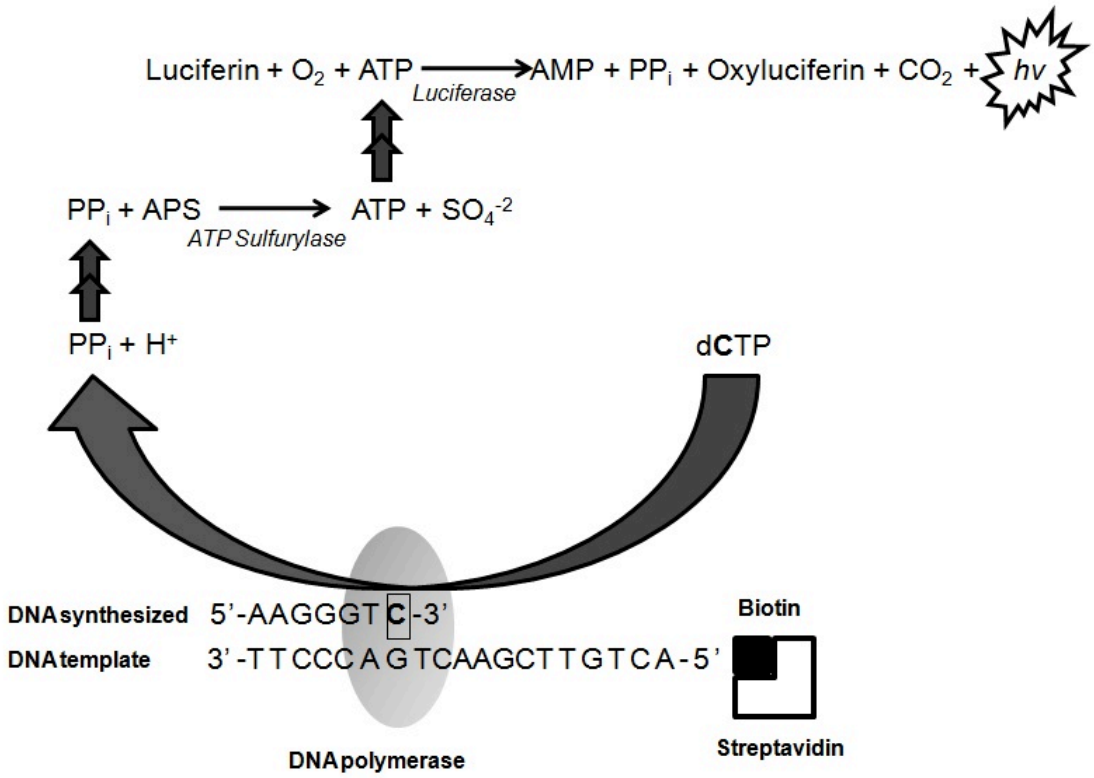
$$PP_i + APS \rightarrow ATP + SO_4^{-2} \text{ (catalyzed by ATP-sulfurylase)} \quad \text{(equation 2)}$$

$$ATP + \text{luciferin} + O_2 \rightarrow AMP + PP_i + \text{oxyluciferin} + CO_2 + h\nu \quad \text{(equation 3)}$$

Thus, light (designated *hv*) emission is proportional to the amount of pyrophosphate produced as a result of the number of nucleotides added. After light is emitted, apyrase catalyzes the degradation of excess dNTPs and ATP via the following reactions.

$$ATP + dNTP \rightarrow AMP + dNMP + 4P_i \quad \text{(equation 4)}$$

Figure 2



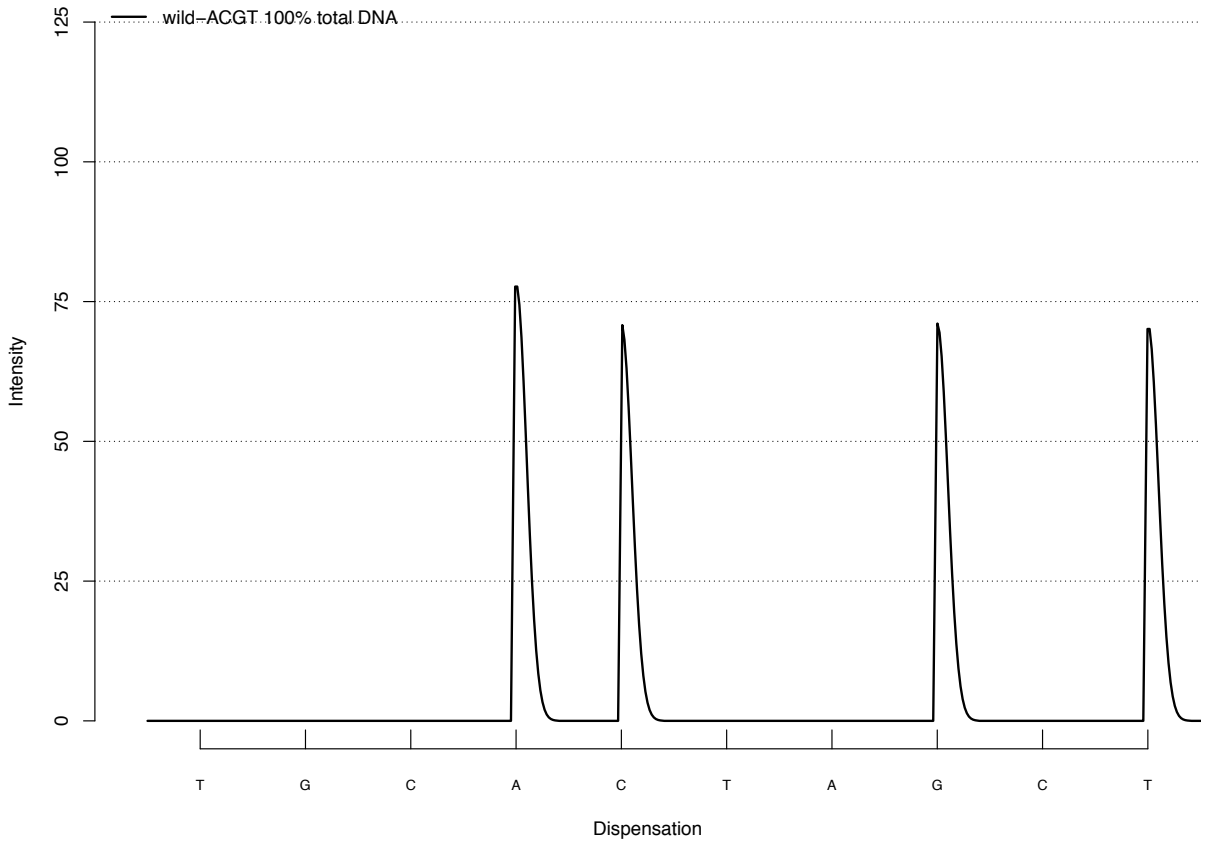
The presence or absence of a light peak following incorporation of a dNTP provides sequencing information

Figure 3: How to generate a pyrogram for a wild type sequence

1. Access Pyromaker at **http://pyromaker.pathology.jhmi.edu/**.
2. Set **Trace Type** to **Pyro**.
3. Enter **limit of detection**. The typical limit of detection for pyrosequencing is **5%**.
4. Enter **0%** for **% Mutation-Containing Cells**.
5. Set **Separate Traces** to **No**.

6. Enter **TGCACTAGCT** for **Dispensation**.
7. Enter **ACGT** into row for **wild** type.
8. Leave the row for **mutant** blank.
9. Leave the box for **Mutant Percent** blank.
10. Set **Chromosome Status** as **Heterozygous**.
11. Click **Generate Graph**.

Figure 3



As expected for the hypothetical wild type sequence $^{5'}$ACGT$^{3'}$ (to be built on a $^{3'}$TGCA$^{5'}$ template strand, not shown), the first three nucleotides dispensed did not extend the growing

strand because the first complementary nucleotide is dATP. Thus, there are no peaks on the pyrogram for the dTTP, dGTP, or dCTP dispensed nucleotide tri-phosphates because they could not be incorporated into the growing strand. However, when dATP was dispensed, it was incorporated, pyrophosphate was released and subsequently, light was emitted. Light is detected by a CCD (charge coupled device) sensor and represented by peaks in the pyrogram.

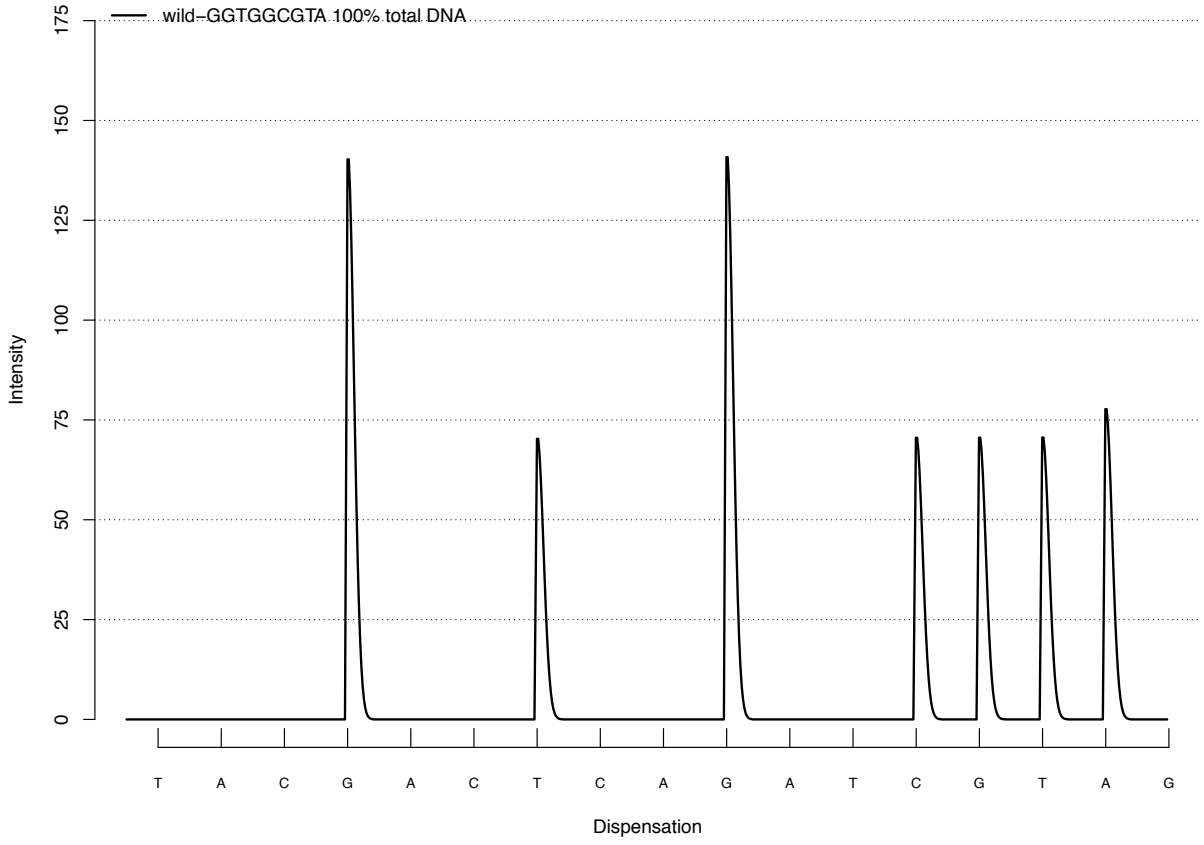
Note that the natural dATP results in false signals because like rATP, it is a substrate for luciferase. Accordingly, the dATP analog, dATP-α-S, is used in lieu of dATP, but produces slightly higher peaks than the other three dNTPs, which should be taken into consideration when comparing homopolymers of A equal in length to the other nucleotides. Notice in this pyrogram how the peaks for a single C, G, and T are of the same height, while the peak for a single A nucleotide is slightly higher.

Peak heights are proportional to the number of nucleotides incorporated

Figure 4: How to generate a pyrogram for wild type KRAS codons 12, 13 and 14

1. Repeat **Steps 1-5** as listed above.
2. Enter **TACGACTCAGATCGTAG** for **Dispensation**.
3. Enter **GGTGGCGTA** into row for **wild** type.
4. Leave the row for **mutant** blank.
5. Leave the box for **Mutant Percent** blank.
6. Set **Chromosome Status** as **Heterozygous**.
7. Click **Generate Graph**.

Figure 4



The height of the peak is generally proportional to the number of identical bases of a homopolymeric run as they are incorporated into the elongating DNA strand. The pyrogram of codons 12, 13 and 14 of KRAS shown above demonstrates the proportionality between peak height and homopolymer length. The first G peak of codon 12 (GGT) is approximately twice as high as the first T peak because the G peak represents two guanines (2X) while the T peak represents only one thymidine nucleotide (1X) in the growing DNA sequence. This is also seen with codon 13 (GGC), but not for codon 14 (GTA) where all nucleotides are present as single bases.

Dispensation Sequence Importance

The dispensation sequence for pyrosequencing can affect how the pyrogram will appear. Different sequences will produce non-identical pyrograms for the same simple mutation, defined as a single base substitution. Two options for dispensation sequences exist: cyclic or optimized programmed sequence unique to the DNA sequence of interest. A programmed sequence is more efficient at detecting mutations and facilitates a longer read length, a faster readout, and less out of phase shifts.

Figure 5a: How to generate a pyrogram with optimized dispensation for wild type and mutant KRAS codons 12, 13 and 14

1. Repeat **Steps 1-5** as listed above.
2. Enter **TACGACTCAGATCGTAG** for **Dispensation**.
3. Enter **GGTGGCGTA** into row for **wild** type.
4. Enter **GATGGCGTA** into row for **mutant**.
5. Set % Mutation containing cells to 50%
6. Set **Mutant Percent** to **100%**.
7. Set **Chromosome Status** as **Heterozygous**.
8. Click **Generate Graph**.

Figure 5a

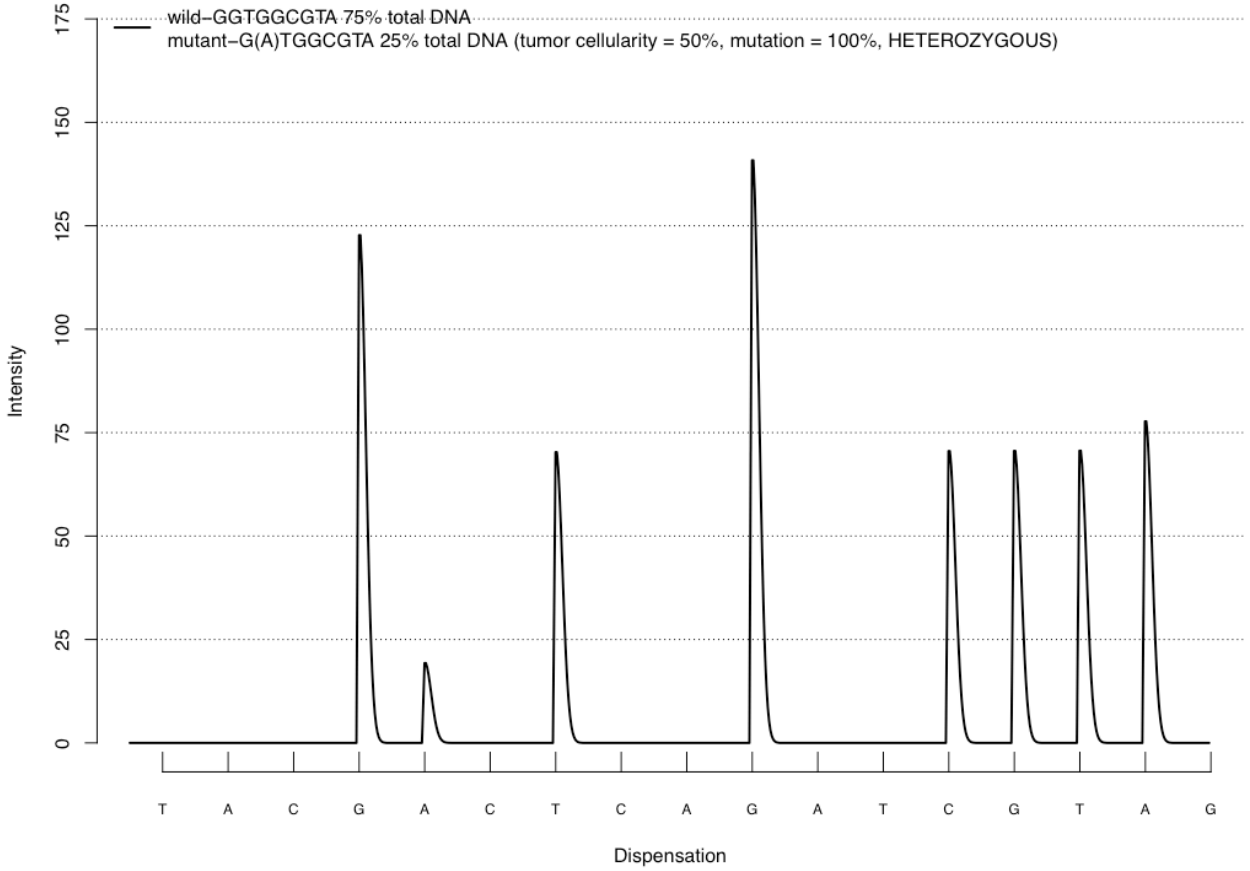


Figure 5a uses a dispensation sequence that is optimized for KRAS. The optimized dispensation sequence generates a cleaner pyrogram with an easily identifiable mutant peak when the second dATP dispensed (Figure 5a).

Figure 5b: How to generate a pyrogram with cyclic dispensation for wild type and mutant KRAS codons 12, 13 and 14

1. Repeat **Steps 1-5** as listed above.
2. Enter **AGCT** for **Dispensation**.

3. Enter **GGTGGCGTA** into row for **wild** type.
4. Enter **GATGGCGTA** into row for **mutant**.
5. Set % Mutation containing cells to 50%
6. Set **Mutant Percent** to **100%**
7. Set **Chromosome Status** as **Heterozygous**.
8. Click **Generate Graph**.

Figure 5b

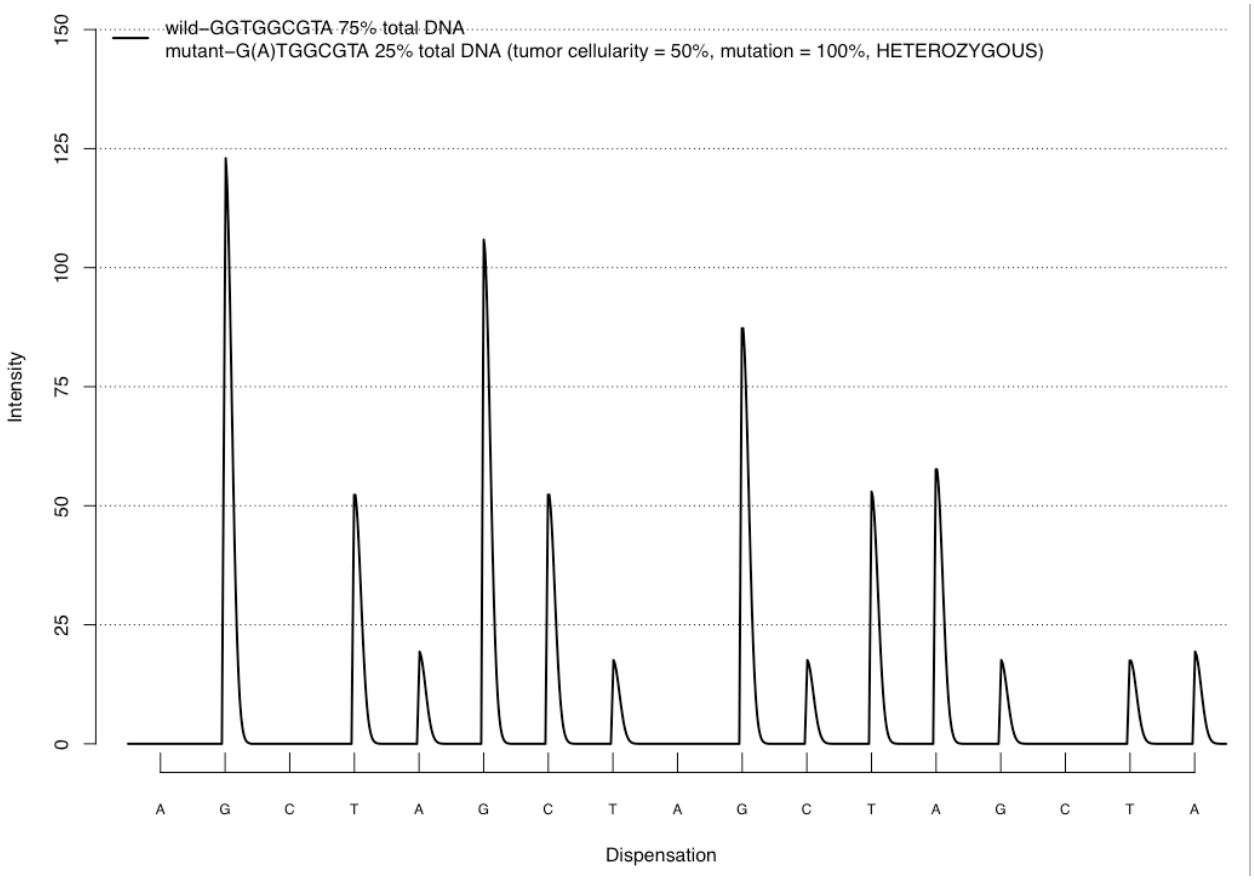


Figure 5b uses a cyclic dispensation sequence (A, G, C, T repeated). In comparing two pyrograms that are detecting the same mutation (5a vs. 5b), note how the two different dispensation sequences create different pyrograms for this simple 12b KRAS mutation (G→A). The cyclic dispensation sequence results in an extremely complex pyrogram with more peaks,

making the pyrogram extremely difficult to analyze. The presence of additional peaks indicates that the wild type and mutant primer strands are out of phase with each other.

In Phase and Out of Phase Pyrograms

Phase is defined as the relative positions of the DNA polymerase molecules on two DNA strands in different sequences and is affected by dispensation order. Elongating DNA strands are in phase when they are aligned at the same base location on both the wild type and the mutant sequences. Nascent strands are out of phase when one is at a further base location than the other strand. When the unknown is out of phase with the wild type, it suggests that the unknown sequence contains a mutation. In contrast to cyclic dispensation, the optimized dispensation sequence is more effective at keeping the mutant, such as 12b KRAS, and wild type samples in phase at the base locations that are identical in both samples while leaving the base locations that are truly mutant ahead of the wildtype sequence.

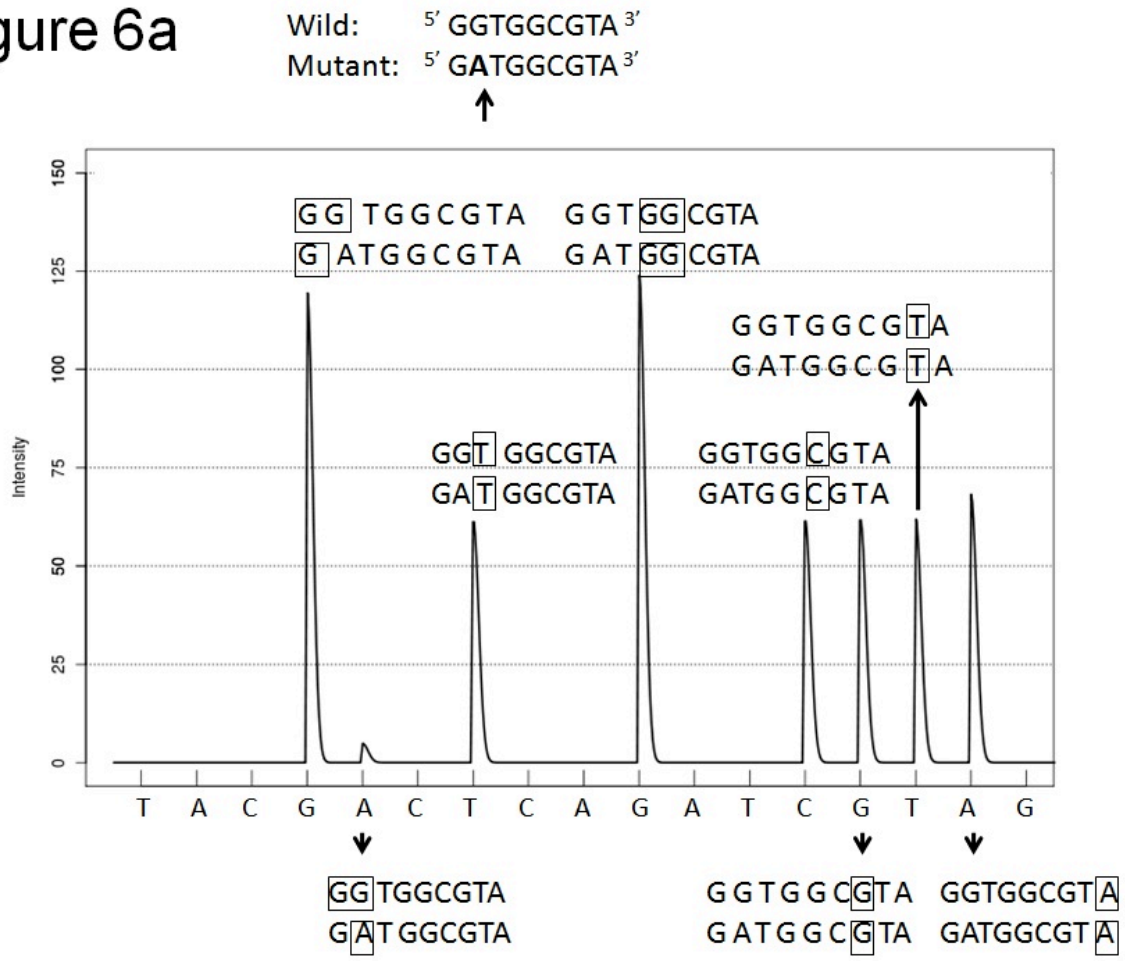
Figure 6a: How to read a pyrogram with optimized dispensation for wild type and mutant KRAS codons 12, 13 and 14

1. Repeat **Steps 1-5** as listed above.
2. Enter **TACGACTCAGATCGTAG** for **Dispensation**.
3. Enter **GGTGGCGTA** into row for **wild** type.
4. Enter **GATGGCGTA** into row for **mutant**.
5. Set % Mutation containing cells to 50%
6. Set **Mutant Percent** to **100%**.

7. Set **Chromosome Status** as **Heterozygous**.

8. Click **Generate Graph**.

Figure 6a

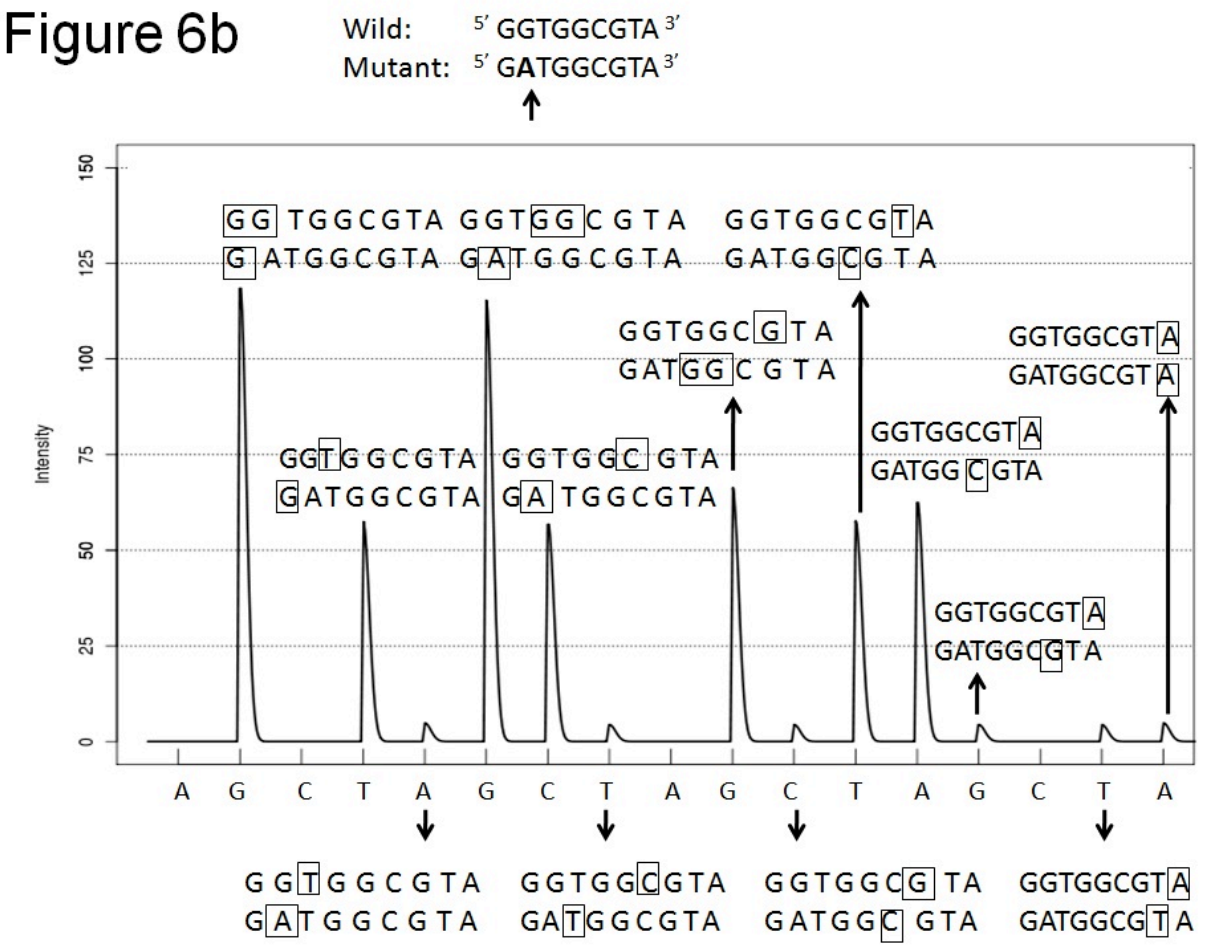


For codons 12, 13, and 14 of wild type KRAS and 12b KRAS mutant (G→A), both alleles are elongated when dGTP is dispensed, but the elongating strands are temporarily out of phase because the wild type allele incorporates two dGTPs while the mutant allele only incorporates one dGTP. The subsequent addition of dATP brings the primers back in phase, as the wild type allele does not advance and the mutant allele incorporates one dATP and advances one nucleotide. Hereafter, the wild type and mutant alleles remain in phase for the rest of the pyrogram. Because the nascent strands are in phase for most of the pyrogram, it is easy to

identify the simple mutation by the peak at the second dATP dispensed. A detailed a base-by-base extension of this pyrogram shows the location of in phase and out of phase primers as sequencing occurs (Figure 7a).

Figure 6b: How to read a pyrogram with cyclic dispensation for wild type and mutant KRAS codons 12, 13 and 14

1. Repeat **Steps 1-5** as listed above.
2. Enter **AGCTAGCTAGCTAGCTA** for **Dispensation**.
3. Enter **GGTGGCGTA** into row for **wild** type.
4. Enter **GATGGCGTA** into row for **mutant**.
5. Set % Mutation containing cells to 50%
6. Set **Mutant Percent** to **100%**.
7. Set **Chromosome Status** as **Heterozygous**.
8. Click **Generate Graph**.



Using an AGCT cyclic dispensation sequence for the identification of the same simple 12b KRAS mutation generates a more complex pyrogram (Figure 5b, 7b). The presence of more peaks in comparison to the pyrogram of the same simple mutation, which uses an optimized dispensation sequence, is due to the wild type and mutant primers continuously being out of phase; because the polymerase is incorporating nucleotides in an unsynchronized manner, many of the peaks represent nucleotides incorporated by either allele but not both. Therefore, out of phase DNA strands can lead to complex pyrograms with more ambiguous peaks. Pyrograms are not optimal for long sequences because wild type and mutant primers will continue to be out of phase as sequence length increases.

A detailed analysis of the pyrogram in Figure 5b depicts a base-by-base reading of the pyrogram, demonstrating how a cyclic dispensation sequence causes primers to stay out of phase (Figure 7b). The cyclic dispensation sequence employed is AGCT. The first nucleotide dispensed, dATP, is not incorporated in either allele, while the next one, dGTP, is incorporated twice on the wild type allele and only once on the mutant allele, putting the strands out of phase. dCTP produces no peak, while dTTP is incorporated by the wild type allele, moving the nascent strand further out of phase with the mutant allele primer by two nucleotides. The second dATP dispensed is incorporated by the mutant allele, yet the strands still remain out of phase. Because the dispensation is cyclic and not optimized, once the primers get out of phase they are unable to get back in phase for the rest of the sequencing. This shows how a cyclic dispensation sequence can generate suboptimal pyrograms for even a simple mutation, especially one located at the beginning of the DNA sequence.

Figure 7: Partitioned pyrogram

Partitioned pyrograms can be useful in understanding which peaks are derived from signal made by sequencing mutant, wildtype or both mutant and wildtype DNA. In this example, one can clearly see the contribution of the wildtype sequence (blue) and the mutant sequence (red) to a peak at a given position.

Figure 7: How to generate a pyrogram for a wild type sequence

1. Access Pyromaker at **http://pyromaker.pathology.jhmi.edu/**.
2. Set **Trace Type** to **Pyro**.
3. Enter **limit of detection**. Minimum for experimental and clinical pyrosequencing is **5%**.

4. Enter 5**0%** for **% Mutation-Containing Cells**.
5. Set **Separate Traces** to ***Yes***.
6. Enter **TGCACTAGCT** for **Dispensation**.
7. Enter **GGTGGCGTA** into row for **wild** type.
8. Enter **GATGGCGTA** into row for **mutant**.
9. Enter 100% for **Mutant Percent**.
10. Set **Chromosome Status** as **Heterozygous**.
11. Click **Generate Graph**.

Figure 7