

From: [Greg Porreca](#)
To: [Young, Dan](#)
Cc: [Waller, Patrick](#)
Subject: Re: G0871.70002US00; Uri's application
Date: Tuesday, May 5, 2009 3:28:05 PM

Hi Dan-

On Tue, May 5, 2009 at 10:30 AM, Daniel Young <Daniel.Young@wolfgreenfield.com> wrote:

Thanks, Greg.

So, $[N! / (N-M)!] / N^M$ is the likelihood of having a set of M target-barcode pairs such that each barcode only appears once in the set. To be sure, is $p=0.05$ the result of $[N! / (N-M)!] / N^M$ based on $N=1E9$ and $M=1E4$? Based on your example it seems that the ploidy of the genome equivalent is haploid. For a diploid genome the calculation would be based on 20000, assuming two copies per target. Is this correct?

The formula is for the probability that there are no repeats. So 1-p would be the 5%. You're correct about haploid/diploid.

The second paragraph of the example below states that the specific number of barcodes necessary for a diploid basecall depends on the level of confidence desired and is calculated from the binomial probability distribution. Could you expand on this concept, perhaps by way of an example.

We're working on this and will have a histogram for inclusion hopefully tomorrow.

Thanks
Dan

From: Greg Porreca [mailto:gporreca@gsgenetics.com]
Sent: Tuesday, May 05, 2009 7:59 AM
To: Young, Daniel W.; Waller, Patrick
Subject: Re: G0871.70002US00; Uri's application

Hi Dan-

We're finished working out Uri's example. The probability paragraph should be as follows:

Given our set of 1000 targets, our task is to determine how many barcodes are necessary to be confident (within some statistical bounds) that a given barcode sequence will not be observed more than once by chance on a given target. For a barcode length of 15 nt, there are 1,073,741,824 possible barcodes present in the MIP reaction, and generally 10,000 genome equivalents in a MIP reaction. The probability of performing a tagging reaction and ending up with one or more copies of a target having the same barcode sequence is calculated as:

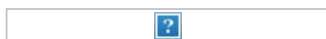
$$p = [N! / (N-M)!] / [N^M]$$

N = # of barcodes, $\sim 1e9$

M = # of genome equivalents, $1e4$

$p = 0.05$

Therefore, our reaction will produce very few (usually 0, but occasionally 1 or more) targets where multiple copies are tagged with the same barcode sequence.



On Thu, Apr 30, 2009 at 8:08 AM, Greg Porreca <gporreca@gsgenetics.com> wrote:

Hi Dan-

We have 2 paragraphs to drop into Uri's application.

The first replaces paragraph 2 of the example (starting with "Given our set of 1000 targets":

Given our set of 1000 targets, our task is to determine how many barcodes are necessary to be confident (within some statistical bounds) that a given barcode sequence will not be observed more than once by chance on a given target. For a barcode length of 10 nt, there are 1,048,576 possible barcodes present in the MIP reaction, and generally 10,000 genome equivalents in a MIP reaction. The probability of tagging two copies of a target with the same barcode sequence, then, is calculated as:

$$\left[\frac{N!}{(N-M)!} \right] / \left[N+M-1 \text{ choose } N-1 \right]$$

N = # of barcodes, ~1e6

M = # of genome equivalents, 1e4

If we sequence our target set to a mean coverage of 100x per target, the probability of observing two actually different capture events with the same target:barcode association is $100 * [XXXX]$, or $[XXXX]$. Therefore, a barcode of length 10 should be more than sufficient for our MIP reaction sequenced to 100x coverage.

Note the two [XXXX] placeholders -- Uri is still calculating the actual numbers, but in the interest of time we decided to get this to you now, and will drop those in when available.

The second goes at the end of the Example, and describes how the barcode count is used during analysis:

A further step that should be taken is to determine whether, after collapsing, there have been enough independent capture events (i.e. there are enough different barcode sequences) to guarantee (for a desired statistical confidence level) that the MIP reaction adequately sampled both alleles in the diploid genome. The specific number of barcodes necessary depends on the level of confidence desired and is calculated from the binomial probability distribution. One can therefore limit analysis to only those positions in the target set which have been captured an adequate number of times to yield an accurate diploid basecall. Therefore, using MIP coverage alone (derived from the bias-detecting barcodes), one will know whether or not a diploid basecall can be made.

On Wed, Apr 29, 2009 at 1:42 PM, Greg Porreca <gporreca@gsgenetics.com> wrote:
sure -- that works for me.

Gregory Porreca, Ph.D.
Director of Technology
Good Start Genetics

8 St Mary's Street, Suite 621
Boston, MA 02215
P: 617-358-6004
F: 617-358-5677

On Wed, Apr 29, 2009 at 1:33 PM, Daniel Young <Daniel.Young@wolfgreenfield.com> wrote:
Greg,

We are up against another meeting. Could we start at 2PM instead?

Thanks
Dan

From: Greg Porreca [mailto:gporreca@gsgenetics.com]
Sent: Wednesday, April 29, 2009 9:56 AM
To: Young, Daniel W.

Subject: Re: G0871.70002US00; Uri's application

Sounds good. Let's plan on my line 617-358-6004, but this may change to our conference line if Uri joins.

-Greg

On Wed, Apr 29, 2009 at 9:52 AM, Daniel Young <Daniel.Young@wolfgreenfield.com> wrote:
Greg,

Let's talk at 1:45. What number should we use to reach you?

Thanks
Dan

From: Greg Porreca [mailto:gporreca@gsgenetics.com]
Sent: Tuesday, April 28, 2009 8:48 PM

To: Young, Daniel W.
Cc: Waller, Patrick
Subject: Re: G0871.70002US00; Uri's application

Hi Dan-
I'm available tomorrow from 1:45 on -- when is best for you?
-Greg

On Tue, Apr 28, 2009 at 5:11 PM, Daniel Young <Daniel.Young@wolfgreenfield.com> wrote:
Greg,

Thanks for your feedback. Please see the attached revised claim set for Uri's application. We would like to discuss these claims with you tomorrow in the context of US Patent Application 2007/0020640, which was noted in the disclosure materials that you sent us. A copy of this application is attached for your reference. Please let me know your availability for a call tomorrow in the early afternoon.

Regards,

Dan

-----Original Message-----

From: Greg Porreca [mailto:gporreca@gsgenetics.com]
Sent: Monday, April 27, 2009 10:08 AM
To: Young, Daniel W.
Cc: Waller, Patrick
Subject: Re: G0871.70000US00; METHOD FOR REDUCING REPRESENTATIONAL BIAS IN MULTIPLEX AMPLIFICATION OF NUCLEIC ACIDS (Porreca)

Hi Dan-

This looks good to me. Uri had one comment:

do you think it makes sense to add additional claims, like "the method of claim X, where the barcodes are random-sequence polynucleotides" or something to that effect?

-Greg

Gregory Porreca, Ph.D.
Director of Technology
Good Start Genetics

8 St Mary's Street, Suite 621
Boston, MA 02215
P: 617-358-6004
F: 617-358-5677

On Fri, Apr 24, 2009 at 7:03 PM, Daniel Young <Daniel.Young@wolfgreenfield.com>
wrote:

Hi Greg,

Please see the attached claims for Uri's application. Please let us know your comments.

Thanks
Dan

-----Original Message-----

From: Young, Daniel W.
Sent: Friday, April 24, 2009 4:38 PM
To: 'Greg Porreca'
Cc: Waller, Patrick
Subject: G0871.70000US00; METHOD FOR REDUCING REPRESENTATIONAL BIAS IN
MULTIPLEX AMPLIFICATION OF NUCLEIC ACIDS (Porreca)

Greg,

Attached is an updated version of the above-mentioned application for you to review. I will forward you a revised set of claims for Uri's application later today. In the meantime, could you send me a document containing the final figures for Uri's application.

Thanks
Dan

Daniel W. Young, Ph.D.
Patent Agent
dwyoung@wolfgreenfield.com
617.646.8245

Wolf Greenfield
Specialists in Intellectual Property Law
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000 | 617.646.8646 fax

For more information about Wolf Greenfield, please visit us at
<http://www.wolfgreenfield.com>

This e-mail message and any attachments may contain confidential or privileged information. If you are not the intended recipient, please notify me immediately by replying to this message. Please destroy all copies of this message and any attachments. Thank you.

