

From: [Waller, Patrick](#)
To: [Greg Porreca](#); [Waller, Patrick](#); [Young, Dan](#)
Cc: [Eric Boutin](#)
Subject: RE: FW: Target Table
Date: Thursday, April 23, 2009 9:04:26 AM

Hi Greg,
I think that you have sent us all of the important information that we need. Dan or I will send you updated drafts today for your comment and feedback. They will include a few questions on issues that we want to clarify, but they should not need significant technical input at this stage.
Patrick

Sent from my Windows Mobile® phone.

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

From: Greg Porreca <gporreca@gsgenetics.com>
Sent: Wednesday, April 22, 2009 2:03 PM
To: Waller, Patrick <Patrick.Waller@wolfgreenfield.com>; Daniel Young <Daniel.Young@wolfgreenfield.com>
Cc: Eric Boutin <eboutin@gsgenetics.com>
Subject: Re: FW: FW: Target Table

Hi Patrick and Dan-

Just checking in on the status of our 3 applications. Please let me know if there's anything I can provide to help move them along.

Thanks-
Greg

On Thu, Apr 16, 2009 at 1:16 PM, Greg Porreca <gporreca@gsgenetics.com> wrote:

> Hi Dan and Patrick-
>
> Attached is my redlined version of the draft. I made the changes (and
> inserted comment balloons) with "Track Changes", so be sure to turn this on
> before viewing.
>
> -Greg
>
>
>
>
>
>
> On Wed, Apr 15, 2009 at 11:15 PM, Daniel Young <
> Daniel.Young@wolfgreenfield.com> wrote:
>
>> Greg,
>>
>> Attached is a draft of the application for you to review. We would like
>> to arrange a time tomorrow to discuss bias tolerance and analytical aspects
>> of the diagnostic methods. Please let us know your availability for a phone
>> call.
>>



ML-IPR-000375

>> Regards,
>>
>> Dan
>>
>>
>>
>> -----Original Message-----
>> *From:* Greg Porreca [<mailto:gporreca@gsgenetics.com>]
>> *Sent:* Wednesday, April 15, 2009 4:19 PM
>> *To:* Young, Daniel W.
>> *Cc:* Eric Boutin; Waller, Patrick
>> *Subject:* Re: FW: FW: Target Table
>>
>> Hi Dan-
>>
>> I've attached an image below that illustrates what bias/efficiency refers
>> to. Each curve is actually thousands of points that naturally form a line
>> because they are sorted along the X-axis by Y-axis value. Each point is the
>> number of times a specific target is observed when the pool of MIP-captured
>> material is sequenced (remember, each of these curves is from approx 5
>> million total observations/sequencing reads split across 13-55K targets by
>> the abundance distribution shown). Bias is the degree to which these curves
>> are not 'flat' (since if the MIP capture reaction were unbiased, all targets
>> would form MIP products at the same rate, and so the abundances of all
>> targets would be the same number). 'Efficiency' is simply the assumed
>> reason for the bias, since MIPs which form either more often than expected
>> (left side of curve) or less often (right side) are assumed to be more or
>> less efficient, respectively.
>>
>> To put numbers to current bias (and this is all target-level bias), from
>> this graph (red and blue curves being state of the art), I'd say most
>> targets (~60%) are within 10-fold of the median efficiency, and ~90% are
>> within 100-fold of the median, with total range of 4 logs.
>>
>> Since most targets are within a 10-fold range, in general averaging across
>> several (n=3 to 10) MIPs should get that set within a factor of 2 the vast
>> majority (e.g. 90%) of the time. For targets out at the extremes, similar n
>> (3-10) should get close to the median (say, within 4-5 fold) but I'd expect
>> more variance. We could probably get hard numbers by simulating, but that
>> will likely take time. If you think it's necessary, let me know and I'll
>> take a crack.
>>
>> For our application where cost is a significant concern, I'd say we'd like
>> 90% of the targets within a 4-5 fold range. But keep in mind that this is a
>> business decision driven by the cost / practicality tradeoff of more
>> sequencing versus more MIPs.
>>
>>
>> [image: Untitled.png]
>>
>> Gregory Porreca, Ph.D.
>> Director of Technology
>> Good Start Genetics
>>
>> 8 St Mary's Street, Suite 621
>> Boston, MA 02215
>> P: 617-358-6004

ML-IPR-000376

>> F: 617-358-5677
>>
>>
>>
>>
>>
>> On Wed, Apr 15, 2009 at 3:28 PM, Daniel Young <
>> Daniel.Young@wolfgreenfield.com> wrote:
>>
>>> Thanks, Greg. Your description does make sense. It is similar, to
>>> some extent, with the "staggered probe layout" of figure 2. Patrick and I
>>> will discuss this.
>>>
>>> We are nearly complete with a draft of this application. It is
>>> important that we are clear on the bias issue, because this is what the
>>> methods seek to minimize.
>>>
>>> What parameter(s) would one use to quantify bias (at the nucleotide
>>> level, the subregion level, the target nucleic acid level, probe set level).
>>>
>>> Could you give us some numbers on the extent to which bias may be
>>> observed without employing the methods (e.g., range, variance), the extent
>>> to which (and nature by which) it may be minimized by employing the
>>> methods (eliminating outliers, reducing variance), and the level of bias (or
>>> type of bias) that may be tolerable in a practical application of the
>>> methods.
>>>
>>> Could you provide us with an equation for 'efficiency' (as mentioned in
>>> your text).
>>>
>>> Thanks
>>>
>>> Thanks
>>> Dan
>>>
>>> -----
>>> *From:* Greg Porreca [<mailto:gporreca@gsgenetics.com>]
>>> *Sent:* Wednesday, April 15, 2009 1:49 PM
>>> *To:* Young, Daniel W.
>>> *Cc:* Eric Boutin; Waller, Patrick
>>>
>>> *Subject:* Re: FW: FW: Target Table
>>>
>>> Hi Dan-
>>>
>>> Sorry, the xls is attached.
>>>
>>> Glad you caught that error on the figure -- new version is attached.
>>>
>>> Your terminology for the claim is fine, but the language about all
>>> subregions being "different" illuminated a hole we may want to consider.
>>> Suppose I'm interested in capturing a 10bp sequence. Rather than either
>>> tiling across or 'staggering' the ends of the subregions, suppose I just
>>> staggered the 'outside' edges of the MIPs such that each MIP had a different
>>> length of sequence hybridizing to the target (but in all cases the
>>> 'captured' sequence between the targeting arms would be the same). For
>>> example, I could have a set of 3 MIPs with targeting arms of length 20, 25,

ML-IPR-000377

>>> and 30. I could potentially still benefit from different targeting arm
>>> hybridization sites (and thus average the performance across partially
>>> different sequences) but capture the exact same subregion every time. Does
>>> this make sense?
>>>
>>>
>>>
>>> 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 15, 2009 at 11:00 AM, Daniel Young <
>>> Daniel.Young@wolfgreenfield.com> wrote:
>>>
>>>> Thanks, Greg.
>>>>
>>>> Thanks for working on the table, could you forward the updated version.
>>>>
>>>> Regarding the figure, the target length (between the two broken lines)
>>>> is 160 bp not 200bp. A 200 bp fragment requires 2 additional probes to get
>>>> 3 times coverage.
>>>>
>>>> A revised claim covering the method for reducing bias in multiplex
>>>> reactions is shown below. Please let us know your thoughts and let us know
>>>> if you agree with the terminology. For example, we have used "subregions of
>>>> the target nucleic acid" rather than "subtargets".
>>>>
>>>> Thanks
>>>> Dan
>>>>
>>>> *Claim:*
>>>>
>>>>
>>>> A method of analyzing a plurality of genetic loci, the method
>>>> comprising:
>>>>
>>>>
>>>>
>>>> contacting each of a plurality of target nucleic acids with a probe set,
>>>>
>>>>
>>>>
>>>> wherein each probe set comprises a plurality of different probes,
>>>> each probe having a central region flanked by a 5' region and a 3' region
>>>> that are complementary to nucleic acids flanking the same strand of one of a
>>>> plurality of subregions of the target nucleic acid, wherein the subregions
>>>> of the target nucleic acid are different, and wherein each subregion

ML-IPR-000378

>>>> overlaps with at least one other subregion,
>>>>
>>>>
>>>> isolating a plurality of nucleic acids each having a nucleic acid
>>>> sequence of a different subregion for each of the plurality of target
>>>> nucleic acids, and
>>>>
>>>>
>>>> analyzing the isolated nucleic acids.
>>>>
>>>>
>>>> -----
>>>> *From:* Greg Porreca [<mailto:gporreca@gsgenetics.com>]
>>>> *Sent:* Wednesday, April 15, 2009 8:42 AM
>>>> *To:* Young, Daniel W.
>>>> *Cc:* Eric Boutin
>>>> *Subject:* Re: FW: FW: Target Table
>>>>
>>>> Hi Dan-
>>>>
>>>> Responses in bold below.
>>>> Please let me know if you need clarification or more info on any of
>>>> these.
>>>>
>>>> Thanks-
>>>> Greg
>>>>
>>>>
>>>>
>>>> On Tue, Apr 14, 2009 at 8:16 PM, Daniel Young <
>>>> Daniel.Young@wolfgreenfield.com> wrote:
>>>>
>>>>> Hi Greg,
>>>>>
>>>>> A few things regarding the application.
>>>>>
>>>>> For several of the genes in the table the RefSeqGene column contains
>>>>> "NM" numbers, which correspond to the transcript sequences. It would be
>>>>> best to have genomic sequences, since this is what will be detected. Could
>>>>> you populate the table in these rows (highlighted in yellow) with genomic
>>>>> REFSEQ IDs. These have "NG" or "NC" numbers.
>>>>>
>>>>>
>>>>> *In cases where the gene is not included in the RefSeqGene list, the
>>>>> only genomic record linked is the reference sequence of the relevant
>>>>> chromosome. I've updated the spreadsheet to reflect this (and found several
>>>>> cases where I missed a RefSeqGene ID that was present).*
>>>>>
>>>>>
>>>>>
>>>>> We would like to describe the relationship between probe layout and
>>>>> region size with reference to specific sizes or ranges of sizes. For
>>>>> example, the disclosure indicates that small regions can be detected by
>>>>> staggered probes and larger regions can be detected with tiled
>>>>> probes. Could you please specify actual region sizes in this context.
>>>>>

ML-IPR-000379

>>>>
>>>> *A single MIP generally can capture a region of length 1bp to 200 bp
>>>> (so in our exon-capture context, most exons would not 'need' to be tiled).
>>>> BUT, you could still imagine a desire to tile, e.g. a case where you
>>>> end-sequence each MIP, rather than prepare a shotgun library, to cover the
>>>> set of exons. In this case the tiling period is driven by the sequencer's
>>>> readlength. The Shendure paper I sent to you and Patrick in separate email
>>>> demonstrates this mode of sequencing (although not this mode of tiling to
>>>> capture/cover large targets).*

>>>>
>>>>
>>>>>>
>>>>>> Could you update the tiled probe layout figure. The figure depict 10
>>>>>> probes. However, based on the configuration described in the text it seems
>>>>>> that there should be 12 probes to cover 200 bp.

>>>>>>
>>>>>>
>>>>>> *To my read, the configuration in the text agrees with the diagram --
>>>>>> what am I missing?*

>>>>>>
>>>>>>
>>>>>>>>
>>>>>>>> We suggest including a generic formula to compute probe number for
>>>>>>>> example based on target length (T), sub-target length (S), and coverage
>>>>>>>> factor (C) (probe number = $T/(S/C) + (C-1)$). Please let us know your
>>>>>>>> thoughts.

>>>>>>>>
>>>>>>>>
>>>>>>>> *That sounds great!*

>>>>>>>>
>>>>>>>>
>>>>>>>>
>>>>>>>>>>
>>>>>>>>>> In addition, we would like to describe the relationships among probe
>>>>>>>>>> number, target length, sub-target length, and coverage factor. Are there
>>>>>>>>>> any specific configurations that are preferred? ranges of sub-target
>>>>>>>>>> lengths? ranges of coverage factors? may sub-target lengths differ within
>>>>>>>>>> a target? Any additional information that you can provide would be helpful.

>>>>>>>>>>
>>>>>>>>>>
>>>>>>>>>> *The coverage factor will be driven by the amount of bias/variability
>>>>>>>>>> you're trying to cancel -- more bias/variability will necessitate more
>>>>>>>>>> subtargets so that the results of the averaging will stay within a narrow
>>>>>>>>>> range.

>>>>>>>>>>
>>>>>>>>>> Sub-target lengths could certainly differ within a target -- for clarity
>>>>>>>>>> I didn't include. You'd do this, for example, if there were 'problematic'
>>>>>>>>>> regions that would break up the 'fixed' tiling period -- you could imagine a
>>>>>>>>>> short region of, say, trinucleotide repeats that you'd want to avoid using
>>>>>>>>>> as a targeting arm site.

>>>>>>>>>>
>>>>>>>>>> The length of the subtarget is driven by different considerations. One
>>>>>>>>>> would be readlength if you were feeding these MIPs directly into a next-gen
>>>>>>>>>> sequencer -- you'd want the subtarget length to match the readlength so
>>>>>>>>>> library construction was not necessary. In general, there isn't a hard and
>>>>>>>>>> fast rule. Your subtarget length could be as short as 1 (where for a
>>>>>>>>>> coverage factor of 3, you'd capture a 100 bp exon as a series of 300
>>>>>>>>>> subtargets), and could 'convert' into the staggered approach as the

ML-IPR-000380

>>>> subtarget length reached the target length.*
>>>>
>>>>
>>>>> Also, please describe the relationship between probe length and
>>>>> sub-target length. Are there preferred sizes? ratios?
>>>>>
>>>>>
>>>> *No real preferred sizes/ratios, except that MIPs (today) are only
>>>> efficient if the captured target is several hundred bp or shorter, so 'long'
>>>> targets necessitate subtargeting at or below this MIP length limit.*
>>>>
>>>>> The disclosure notes that "the amount of bias-tolerance for
>>>>> pathological MIPs can be 'dialed in' by changing the number of different
>>>>> MIPs used to capture a given target". Could you please define
>>>>> 'bias-tolerance'.
>>>>>
>>>>>
>>>>> *The idea is that different MIPs will have different performance
>>>>> characteristics. The efficiency of certain targets will vary more from one
>>>>> sample to another than other targets. Additionally, the 'nominal'
>>>>> efficiency of different targets will vary -- some will tend to be very
>>>>> efficient, some will exhibit 'average' efficiency, and some will be very
>>>>> inefficient. Potentially different numbers of MIPs will be used for each of
>>>>> these efficiency/variability classes. Bias tolerance is just the amount of
>>>>> variability/bias one is willing to accept **after** having performed the
>>>>> protocols in this invention. There is a tradeoff between
>>>>> economics/practicality and performance; very large numbers of MIPs per
>>>>> target should effectively reduce bias/variability to zero, but at a large
>>>>> cost.*
>>>>>
>>>>>
>>>>> Please let me know if you'd like to discuss any of this.
>>>>>
>>>>> Regards,
>>>>>
>>>>> Dan
>>>>>
>>>>> -----
>>>>> *From:* Greg Porreca [<mailto:gporreca@gsgenetics.com>]
>>>>> *Sent:* Friday, April 10, 2009 3:53 PM
>>>>> *To:* Young, Daniel W.
>>>>> *Cc:* Waller, Patrick
>>>>> *Subject:* Re: FW: Target Table
>>>>>
>>>>> Sorry, Dan. That was the new version of excel.
>>>>> XLS is attached.
>>>>>
>>>>> -Greg
>>>>>
>>>>>
>>>>> On Fri, Apr 10, 2009 at 3:34 PM, Daniel Young <
>>>>> Daniel.Young@wolfgreenfield.com> wrote:
>>>>>

ML-IPR-000381

>>>>> Greg,
>>>>>
>>>>> I can't seem to open the target gene table file. Could you send it in
>>>>> another format (e.g., tab-delimited text file, excel file, word doc).
>>>>>
>>>>> Thanks
>>>>> Dan
>>>>>
>>>>> -----
>>>>> *From:* Greg Porreca [<mailto:gporreca@gsgenetics.com>]
>>>>> *Sent:* Thursday, April 09, 2009 1:45 PM
>>>>> *To:* Waller, Patrick
>>>>> *Cc:* eboutin@gsgenetics.com; Young, Daniel W.
>>>>> *Subject:* Re: FW: Target Table
>>>>>
>>>>> Patrick-
>>>>>
>>>>> Attached is a table of target genes, as well as SNPs for a few of the
>>>>> genes on the list.
>>>>>
>>>>> -Greg
>>>>>
>>>>>
>>>>> On Wed, Apr 8, 2009 at 8:26 AM, Patrick Waller <
>>>>> Patrick.Waller@wolfgreenfield.com> wrote:
>>>>>
>>>>> Eric and Greg,
>>>>> Here is the type of information that would be helpful for the target
>>>>> table that we discussed. As Dan mentions in his e-mail below, any
>>>>> information on particular SNPs that you are targeting would be helpful too.
>>>>> The purpose of this information is to support broad target detection
>>>>> and also provide basis for narrow claims to specific markers or genes that
>>>>> may be used in some of the tests. Providing the information in the table
>>>>> format is the easiest way to input a significant amount of information that
>>>>> can be used in several or most of your applications related to marker
>>>>> detection.
>>>>> So far, we have focussed on the first two applications and will send
>>>>> you updated drafts to you today along with comments, modified claims, and
>>>>> suggested sections for the beginning of the summaries and detailed
>>>>> descriptions to frame the invention and highlight the distinctions over
>>>>> existing technology.
>>>>> I also will follow up on the third application (Uri's), but may not
>>>>> be able to get comments out to you until late this evening or tomorrow
>>>>> morning.
>>>>> Patrick
>>>>>
>>>>> _____
>>>>>
>>>>> From: Young, Daniel W.
>>>>> Sent: Tue 4/7/2009 7:20 PM
>>>>> To: Waller, Patrick
>>>>> Subject: Target Table
>>>>>
>>>>>
>>>>> Patrick,
>>>>>
>>>>> I have attached a target gene table for the GSG team to fill out.

ML-IPR-000382

