Design of Optimal Depth of Coverage

Step1. Parameter setting

You choose “Calling Error Probability”, “Significance Level”, “Pool Size”, “♯carriers of variant”, “Desired Power”, and “Max Depth” to evaluate optimal depth of coverage.

Step2. Run

After setting all the parameters, click on the “Optimal Depth Run” button below the parameter input form.

Powers for each depth in the range of 1 – “Max Depth” are then evaluated. Optimal depth is the minimum depth exceeding the set desired power. The optimal depth is displayed in the text area below the run button together with a plot of power against depth. The desired power is drawn as horizontal line on the plot. The “Clear” button is used to be clear the output text area.

The powers are evaluated based on binomial distribution; therefore the picture is not always drawn as continuous curve. NDesign makes use of Highcharts JavaScript library to draw the picture. A tooltip text with information on each point (depth and power) can be displayed on hovering the plot.

If optimal depth of coverage can not be found in the range (1 – “Max Depth”), that is, all the powers are below the desired power, much larger value for “Max Depth” may be suiatble to be set. However, it is not feasible too large value (e.g., more than 100 “Max Depth”) because it is computational burden for the current JavaScript version of NDesign.

Design of Experiment

Step1. Total sequence

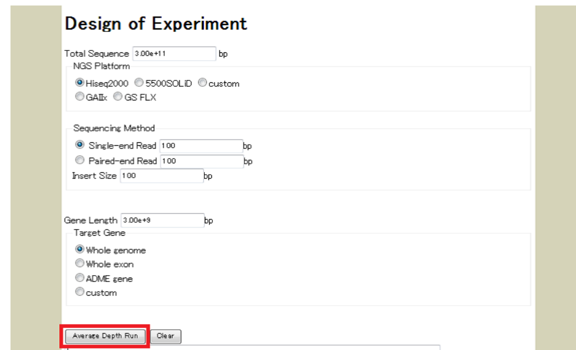
Total sequence is determined by employed NGS platform and sequencing method. At first, you choose the type of NGS platform. The current version of NDesign provides four well-known commercially available NGS platforms: Hiseq2000, GAIIx, 5500SOLiD, and GS FLX. If the sequencing method is selected together with one of the platforms, then NDesign assigns the value for read length as well as for total sequence automatically. You can also set any parameters you prefer. In such a case, you select “custom” option for the selected NGS platform and input the total sequence manually.

Step2. Gene Length

The target gene length of the study may be selected as the next step. NDesign provides options for length of whole genome, whole exon, and principal ADME genes for pharmacogenomics (PGx) as default. Details of principal ADME genes are summarized as Supplementary Table 1 at the last part of this user’s guide. You can also set any gene length after the “custom” option is selected.

Step3. Run

Click on the "Average Depth Run" button below the parameter input forms. The estimated average depth of coverage is then calculated and immediately displayed in the text area below the run button. The "Clear" button is used to be clear the output text area.



The evaluated average depth is also compared with the obtained optimal depth. If the average depth is greater than the optimal depth, the designed experiment may be feasible to be conducted to obtain desired power.

Example usage

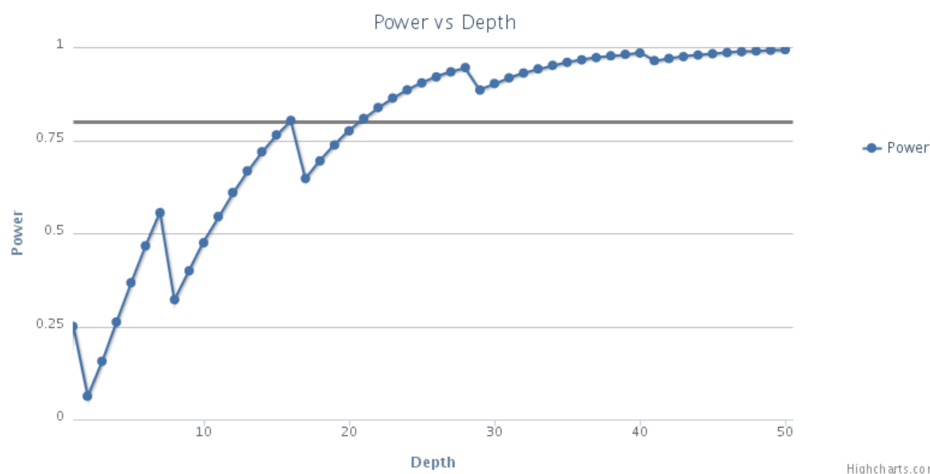
Scenario A (Design of target)

We want to design an experiment to identify rare variants contributing to a certain disease. We have already owned Hiseq2000 and sampled 10 individuals out of which 5 are patients. How do we determine target to obtain desired power?

1. Design of Optimal Depth of Coverage

   We assume that calling error probability, significance level, and desired power are 0.05, 0.05, and 0.8, respectively. Under these assumptions, the optimal depth is obtained as 16X.

   Parameters
   Calling Error Probability : 0.05
   Significance Level : 0.05
   Pool Size : 10 (♯ carriers of variant 5)
   Desired Power : 0.8
   Max Depth : 50



2. Design of Experiment

   We will obtain 10X as the average depth if whole genome is sequenced by single–end read. This experiment is not feasible to be conducted and the target should be changed. The experiment of which target is whole exon or ADME gene is feasible since its average depth is 526X or 14778X.

   Parameters
   NGS Platform : Hiseq2000
   Sequencing Method : Single-end Read (100)
   Target Gene : Whole genome (3.00e+9), Whole exon (5.70e+7), and ADME gene (2.03e+6)

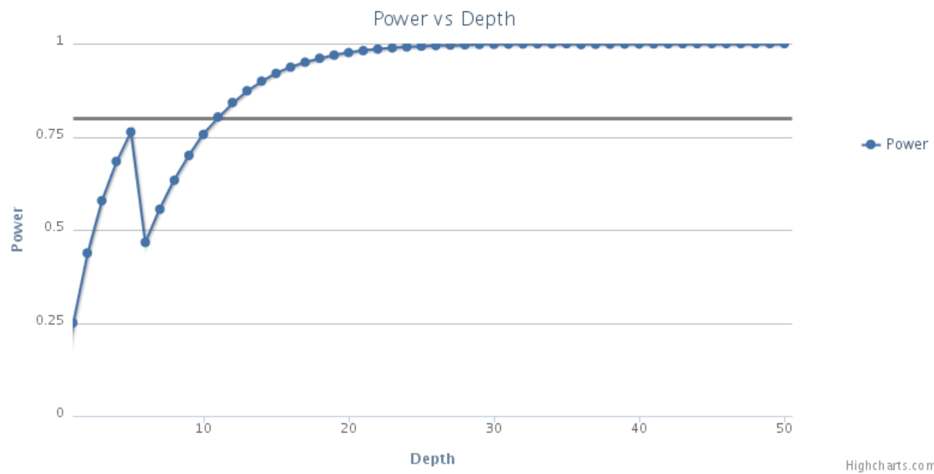
Scenario B (Evaluation of the obtained variants)

We have conducted an experiment and found several candidate variants. Which variant is more suspicious?

1. Design of Optimal Depth of Coverage

   By setting the parameters as follows, the optimal depth is obtained as 11X. The variant locus having more than 11X depth is suspicious.

   <u>Parameters</u>
   Calling Error Probability : 0.01
   Significance Level : 0.05
   Pool Size : 4 (♯ carriers of variant 2)
   Desired Power : 0.8
   Max Depth : 50



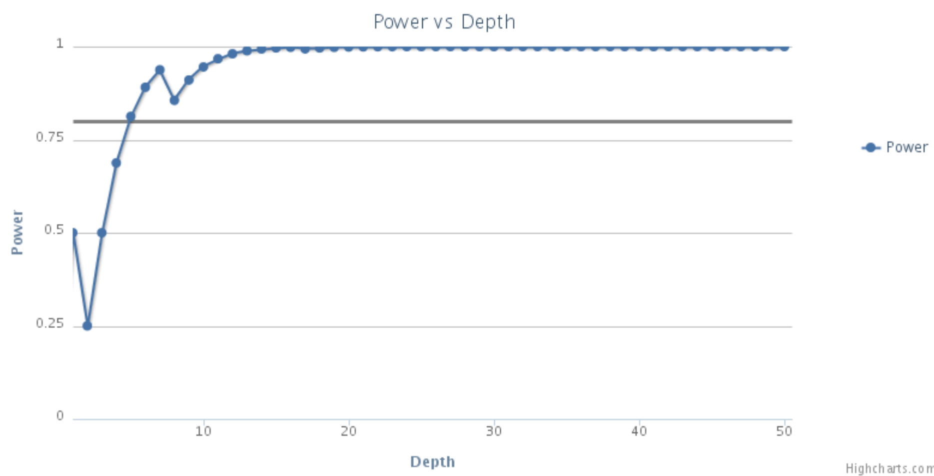
Scenario C (Design of additional experiment)

We have conducted an experiment, but the obtained depths seemed to be low. It was sequenced for an individual and the minimum depth was 2X. How many additional experiments should be conducted to obtain sufficient depth?

1. Design of Optimal Depth of Coverage

   By setting the parameters as follows, the optimal depth is obtained as 5X. It is required 2.5 times of total sequence to obtain the optimal depth for the minimum depth. Therefore, 3 additional runs of experiment should be conducted.

   Parameters
   Calling Error Probability : 0.05
   Significance Level : 0.05
   Pool Size : 1 (♯ carriers of variant 1)
   Desired Power : 0.8
   Max Depth : 50



(APPENDIX)

Supplementary Table 1: Summary of principal ADME genes.

Gene	CHR	Physical position*		Length
		Start	End	
DPYD	1	97543299	98386615	843316
GSTM1	1	110230418	110236367	5949
UGT1A1	2	234668919	234681945	13026
SLC15A2	3	121613171	121663034	49863
UGT2B17	4	69402902	69434245	31343
UGT2B15	4	69512315	69536374	24059
UGT2B7	4	69962193	69978705	16512
ABCG2	4	89011416	89080011	68595
TPMT	6	18128542	18155374	26832
SLC22A1	6	160542863	160579750	36887
SLC22A2	6	160637794	160679963	42169
ABCB1	7	87132948	87342564	209616
CYP3A5	7	99245813	99277621	31808
CYP3A4	7	99354604	99381808	27204
NAT1	8	18067290	18081198	13908
NAT2	8	18248755	18258723	9968
CYP2C19	10	96522463	96612671	90208
CYP2C9	10	96698415	96749148	50733
CYP2C8	10	96796529	96829254	32725
ABCC2	10	101542463	101611662	69199
CYP2E1	10	135340867	135352620	11753
SLC22A6	11	62744069	62752469	8400
GSTP1	11	67351066	67354124	3058
SLCO1B3	12	20963638	21069658	106020
SLCO1B1	12	21284128	21392730	108602
CYP1A1	15	75011883	75017877	5994
CYP1A2	15	75041184	75048941	7757
SULT1A1	16	28616913	28634866	17953
CYP2A6	19	41349443	41356352	6909
CYP2B6	19	41497204	41524301	27097
GSTT1	22	24376139	24384284	8145
CYP2D6	22	42522501	42526883	4382

*The coordinates refer to GRCh37.