xGen™ Hyb Panel Design Tool

Application insight

Hybridization capture is a method of targeted next generation sequencing (NGS). Before hybridization capture is performed, DNA or RNA samples are converted into sequencing libraries. To create libraries, the input material is randomly sheared into smaller fragments, and sequencing adapters are ligated. Depending on the library design, PCR amplification may be required.

Regions of interest within the library are then captured using biotinylated oligonucleotide baits (probes). Probes are designed to hybridize to regions of interest using stocked (e.g., Exome Hyb Panel v2) or custom panels designed to capture targets of specific interest to the user. After hybridization of baits to the fragmented DNA, streptavidin is used to enrich for captured fragments (library molecules bound to baits), while unbound fragments are washed away. Because the DNA is randomly sheared during library preparation, the genomic footprints of captured fragments are overlapping and unique.

Hybridization capture works well for genotyping and rare-variant detection. Since capture with hybridization probes does not require PCR primer design, it is less likely to miss mutations and performs better with respect to sequence complexity. Hybridization capture's capacity for mutation discovery makes it particularly suited to cancer research. The sequence complexity and scalability make it a better choice for exome sequencing than other NGS methods.

xGen hybridization capture solutions support high on-target rates, uniform coverage, and reliable capture performance across a broad range of panel sizes. An updated, automation-friendly protocol enables reproducible, targeted sequencing results across various levels of throughput.

The xGen Hyb Panel Design Tool can assist you in designing your unique panel and empowers you to have those designs in hand quickly. Once your design is ready, order your panels and begin your research. The panels you design can also be queried or viewed at any time.

xGen design process overview

Design of hybridization capture probes using the xGen Hyb Panel Design Tool results in a panel with functional advantages. Upfront input validation ensures probes can be designed using the selected parameters. The design process is different, depending on the type of input submitted.

For BED file designs, targets are merged within a predetermined genomic distance, and a flanking genomic sequence is included as part of the design space. Gene symbol designs are created by first building a BED file representing either the exons belonging to each transcript for that gene, or the entire footprint of the gene from its first to its last exon. The design then proceeds as for BED file designs. The genomic annotation is current as of the NCBI RefSeq release noted for each species in the tool's drop-down selection box.

Note: Some gene symbols have been mapped by NCBI RefSeq both to canonical chromosomes, (e.g., chr6), and alternate contigs, (e.g., chr6_GL000250v2_alt). To avoid generating duplicate probe sequences, the design tool will design probes using only the canonical set of chromosomes for each species.

In contrast to the two BED coordinate-based designs described above, the two other input types (FASTA sequences and accession numbers) are generated directly from target sequences. Transcript accession numbers and versions are current as of the date corresponding to the NCBI RefSeq release noted for each species.

In each case, the probes are selected based on on-target efficiency and specificity using IDT's proprietary probe design process to generate an NGS panel with broad, consistent content coverage.

Glossary

Term	Definition
Panel name	A user-defined name associated with each panel.
Target name	A user- and design tool-defined name associated with a specific target in a panel.
BED format	A text file format used to define genomic regions by coordinates in a chromosome or FASTA sequence. At minimum, a chromosome/sequence name, start, and stop coordinate must be supplied for each record.
FASTA sequence	A text-based format showing a nucleotide (or protein) sequence.
Off-target probe QC	A proprietary off-target prediction method that provides a risk categorization for all probes in the panel.
Tiling	Tiling refers to the number of times, on average, a base in your target is covered by a probe. For example, in 1X tiling designs to BED targets, each base is covered by one probe (probes are designed end to end), whereas in 2X tiling designs, each base is covered by two probes (probes are designed with 40-80 bp overlap).
Gene symbol	A unique abbreviation for a gene name.
Full Region	Selecting this option designs probes to the full gene, including coding exons, introns, and UTRs. This option is only available with gene symbol designs.
Exons (with UTRs)	Selecting this option designs probes to coding exons and UTRs. This option is only available with gene symbol designs.
Exons (without UTRs)	Selecting this option designs probes to coding exons solely. This option is only available with gene symbol designs.
Accession number	A unique identifier for a sequence that is continually updated and includes version history information.

Access the xGen Hyb Panel Design tool

Open the xGen Hyb Panel Design Tool at **www.idtdna.com/xGENdesigntool**.

You will need to log into your IDT web account using your username and password to access the tool.

Create new design

Input format

Select one of the following:

- BED target coordinates
- FASTA target sequence
- Gene symbol
- Accession number

Target definition (only applicable to designs with gene symbol inputs)

Select one of the following:

- Full Region
- Exons (with UTR)
- Exons (without UTR)

Species

The xGen Custom Design Tool uses reference genomes obtained from NCBI to perform predictive, off-target QC on the designed probes, ensuring high specificity.

Select from one of the following species profiles:

Species	Reference genome
Homo sapiens (human)	GRCh37 (hg19)
Homo sapiens (human)	GRCh38 (hg38)
Mus musculus (mouse)	GRCm38 (mm10)
Mus musculus (mouse)	GRCm39 (mm39)
Rattus norvegicus (rat)	Rnor_6.0 (rn6)
Rattus norvegicus (rat)	mRatBN7.2 (rn7)
Canis lupis familiaris	CanFam3.1 (boxer)
Canis lupis familiaris	ROS_Cfam_1.0 (lab)

Tip: If you would like to submit a set of targets for a species that is not found in the dropdown menu, select **Other**. When selecting **Other**, you will be asked to submit FASTA sequences and understand that no off-target QC will be performed. Alternatively, you can reach out to the NGS Design team with a custom design request **here**.

Run off-target probe QC

Predictive off-target QC is automatically included for designs using gene symbols or BED coordinates as inputs and is performed against the genome selected in the species dropdown menu. Predictive off-target QC is optional for inputs provided in FASTA format or as accession numbers.

Note NGS panels designed without predictive off-target QC analysis are not eligible for NGS functional testing and could impact the product options available to you.

Probe length

Probe length is required to be 120 bases.

Note: If you would like probes of a different length, submit a custom design request here.

Probe tiling density

Tiling density can be either 1X or 2X.

Note: If you would like a different tiling strategy, submit a custom design request here.

Provide target inputs

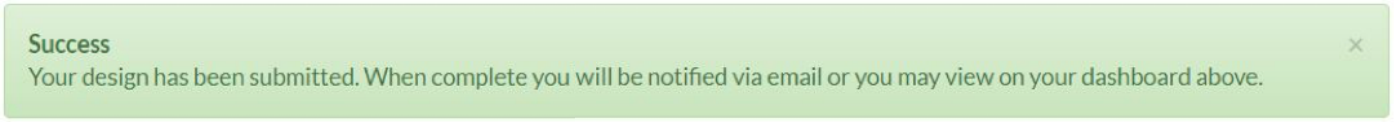
Enter your target input by typing/pasting the target information into the space provided, or upload a file in the **Upload file** tab.

Submit for design

Click DESIGN to submit your request.

The xGen Design Tool has a feature that checks target inputs to maximize the design success rate.

The input QC check should take a few minutes or less to complete. If you do not have any input errors, you will see the message below, indicating that your inputs have been successfully submitted and your design is in process.



Once your designed panel selection is ready, you will receive an email. The time required to complete a design with our QC pipeline depends on your panel size—small panels may take up to an hour, while larger panels may take up to 2 days.

Input QC check

The input QC check assesses correct formatting of inputs, while conforming to the xGen pipeline design rules. If the target inputs do not process immediately upon clicking **Design**, you may see an error like this:

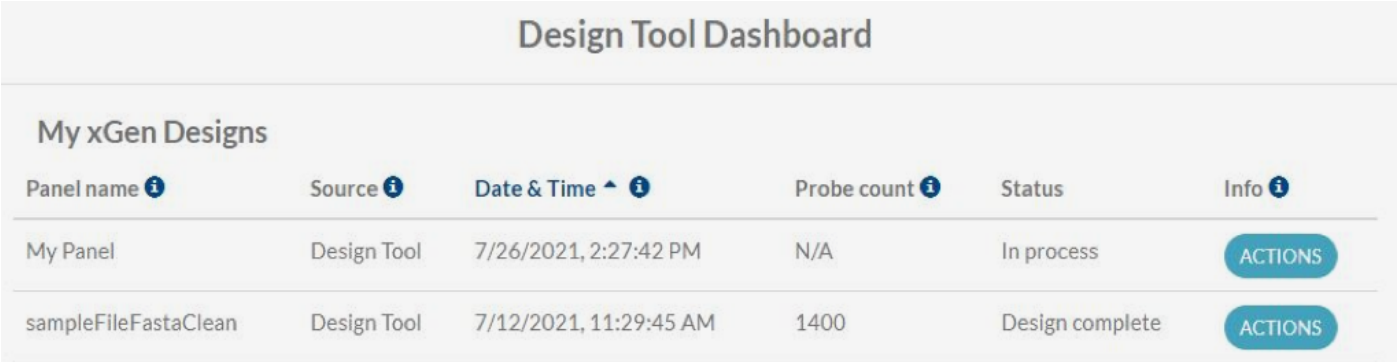
Review the descriptions associated with each target.

You will have the option to **1**) Download the Excel file containing the list of your targets and the QC message associated with each target, **2**) Cancel the design submission and fix the issues with the relevant targets, and then re-submit the corrected inputs, or **3**) Continue with the design submission for all remaining valid targets.

Note Some of the issues identified during input QC check may not need to be corrected, e.g., the merging of two targets in close proximity to each other. In such cases, click **Continue** to submit the design.

Review design and request quote

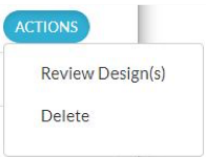
To review your design, go to the xGen Hyb Panel Design Tool web page and click on the **Get started** button. All the NGS panels you have recently created will be listed. If you have requested a custom design by the IDT NGS Design Team, it will also be present in your dashboard.

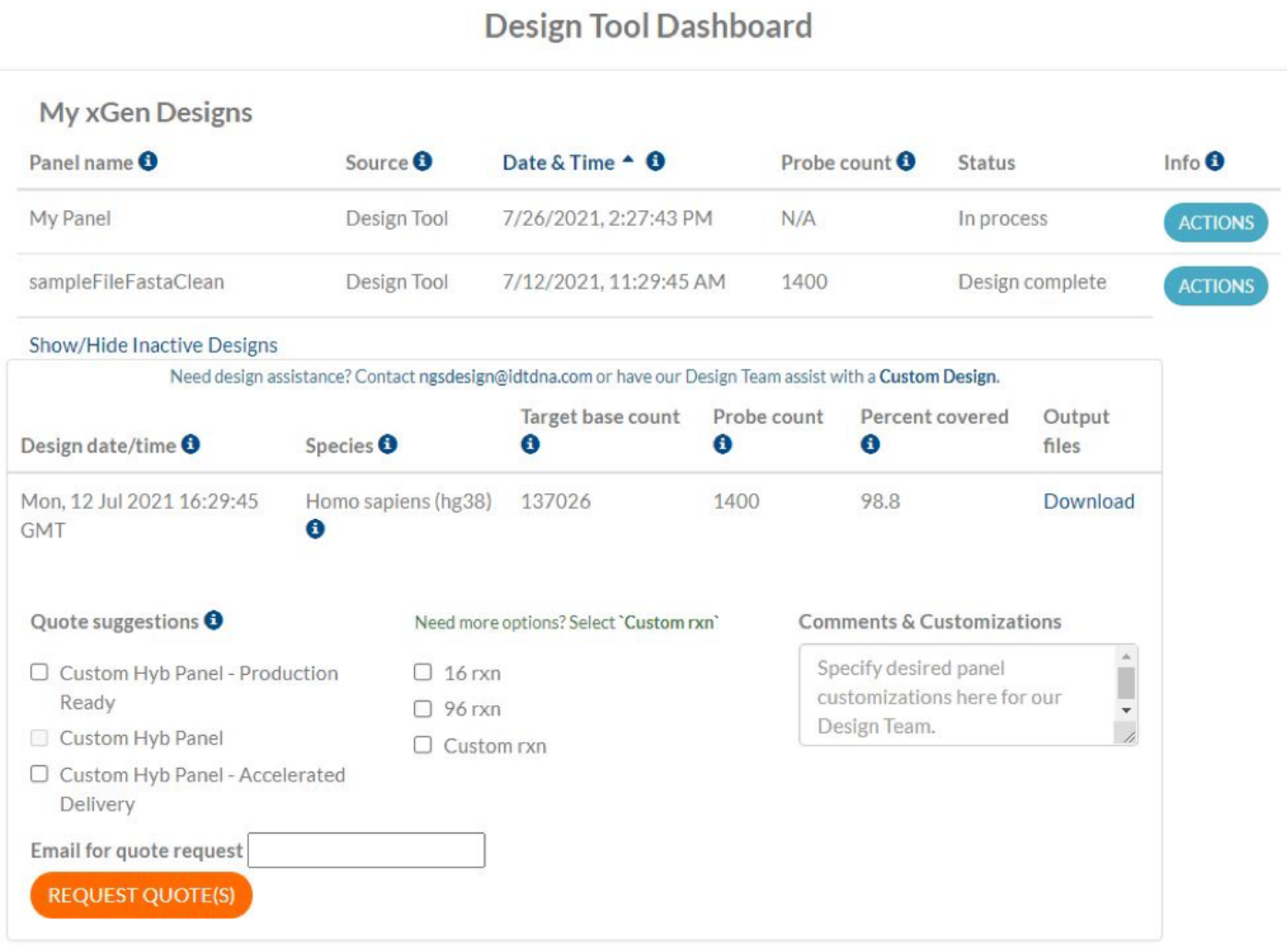


Designs that are complete and ready for review will be indicated by a Status indication of "Design complete." Designs that are still "in process" are not able to be viewed. Designs that have been quoted or ordered will show a status of "Quote requested" and "Ordered", respectively. If you have requested a custom design by a member of the IDT NGS design team, the Status will indicate "Ready for review" when you are able to view the design.

Source refers to the origin of the probe sequences and associated design files which can be "IDT Quote," "IDT Design," or "Design Tool," depending on how the design was initiated. More information on these categories can be found by clicking the ℹ icon next to the Source column.

To review the design files, click on the *ACTIONS* button for that design, which opens two options. Selecting "Review Design(s)" allows design files to be downloaded and quotes requested. Selecting "Delete" removes the design from your dashboard; we only recommend doing this if you no longer need the design and do not plan to order it.





To download design files, click **Download** under the *Output files* column.

Design metadata include information on the species used in the design, the target base count, the probe count, and the percentage of the submitted target regions overlapped by or within 30 bp of the probes.

To request a quote, select the product type and reaction size. You may have multiple product types or reaction sizes quoted at the same time. If you have custom panel formulation requirements, select "Custom rxn" and specify the requirements in the "Comments and Customizations" text box. You may request a different email address for the quote if you require someone other than yourself to be on the quote (e.g., you have a central purchaser) by entering an email address other than your own in the "Email for quote request" text box. Once you have your quote requirements finalized, click on the REQUEST QUOTE(S) button, and a quote will be sent to you by the next business day.

To exit the design and return to the overall dashboard view, click on the **Actions** button again and select "Collapse Design(s)."

Files available to download

- Excel file containing probe sequence and coordinates (probe off-target information included if applicable)
- Target BED file
- Probe BED file

Troubleshooting

BED QC Errors

Message	Suggested action	Notes
Adjacent targets are merged, but the merged target exceeds the size limitation.	Individual designs are limited in size. You may split the design into two smaller designs and try again. Alternatively, you may reach out to Application Support (applicationsupport@idtdna.com) for help.	Design will be unable to proceed until this error is addressed.
BED file input missing 1 or more fields. Chrom, chromStart, and chromEnd are required columns.	Check that the BED file contains at least three fields (chromosome, start coordinate, stop coordinate).	Design will be unable to proceed until this error is addressed.
Chromosome or scaffold name is not found in reference genome.	Confirm that targets correspond to the reference genome assembly selected for the design.	The design can proceed without intervention, but the invalid target will be skipped.
The start/stop coordinate must be >= 0.	Check that the BED file contains at least three fields (chromosome, start coordinate, stop coordinate). All coordinates must be non-negative.	Design will be unable to proceed until this error is addressed.
The start coordinate must be smaller than the stop coordinate.	Ensure that the value in column 3 is higher than the value in column 2.	Design will be unable to proceed until this error is addressed.
The coordinates supplied are beyond the range of the specified chromosome or scaffold.	The coordinates referenced in the BED file fall outside the range of the chromosome found in the reference; confirm that targets correspond to the reference genome assembly selected for the design.	The design can proceed without intervention, but the invalid target will be skipped.
The coordinates of the BED record are a duplicate from another record; the duplicate has been removed from the design.	No action needed, unless you want to remove duplicate BED coordinate entries from the input file.	The design can proceed without intervention.
Duplicated target names have been altered to create a unique ID.	No action needed, unless you want to create your own unique IDs. If any targets have the same name, the Design Tool will append the name to make it unique.	The design can proceed without intervention.
Since no valid strand was specified, the strand has been defaulted to (+).	If you would like a strand-specific design, update the BED file to indicate either (+) or (-) strand orientation.	The design can proceed without intervention.

FASTA QC Errors

Message	Suggested action	Notes
FASTA sequence length must be between 120 and 2500000 characters.	Supplied FASTA sequences must be at least the length of a single probe. There is also an upper limit of 2.5M bp for design tool submissions.	The FASTA entry must be deleted or amended in the FASTA editor before design can proceed.
FASTA sequences can only contain A, C, T, and G.	Bases other than A, C, G, or T are not allowed in probe sequences. The sequence needs to be removed or updated to exclude degenerate bases or lowercase letters.	The FASTA entry must be deleted or amended in the FASTA editor before design can proceed.
Duplicated target names have been altered to create a unique ID.	No action needed, unless you want to create your own unique IDs. If any targets have the same name, the Design Tool will append the name to make it unique.	The design can proceed without intervention.
FASTA names may only contain alpha, numeric, period, dash, and underscore characters.	Supplied FASTA names are restricted to the characters mentioned. Update the names of the sequences accordingly and resubmit.	Design will be unable to proceed until this error is addressed.

GENE SYMBOL Errors

Message	Suggested action	Notes
This gene symbol is not found in the database.	The target has been removed. Double-check the gene symbol to ensure there are no typos in it.	The design can proceed without intervention, but the invalid target will be skipped.
No CDS regions are found for this gene in the database.	The target has been removed. Double-check that the gene has annotated coding exons in the reference genome assembly you selected.	The design can proceed without intervention, but the invalid target will be skipped.
No exons are found for this gene in the database.	The target has been removed. Double-check that the gene has annotated exons in the reference genome assembly you selected.	The design can proceed without intervention, but the invalid target will be skipped.
XXXX1 unambiguously resolved to YYYY1:(GeneID). YYYY1 will be the gene symbol used in the design.	Cases where unofficial gene aliases can be resolved to the official NCBI gene symbol are noted.	The design can proceed without intervention.
XXXX1 is an official gene symbol and will be designed. XXXX1 can resolve to YYYY1: (GeneID). If you want YYYY1, please replace XXXX1 with YYYY1.	Cases where official NCBI gene symbols have also been identified as unofficial gene aliases will be noted. If the resolved alias is your intended target, replace the first symbol in your design set with the second.	The design can proceed without intervention. Only the originally entered gene will receive a design.

GENE SYMBOL Errors (continued)

Message	Suggested action	Notes
XXXX1 is not an official gene symbol. XXXX1 can resolve to YYYY1:(GeneID). If you want YYYY1, please replace XXXX1 with YYYY1. XXXX1 can resolve to ZZZZ1:(GeneID). If you want ZZZZ1, please replace XXXX1 with ZZZZ1.	Cases where unofficial gene aliases resolve to two or more official NCBI gene symbols are noted. Replace the original symbol in your design with the official gene symbol(s) you intend to target.	The design can proceed without intervention, but no design will be created for the given gene symbol.
This gene symbol is duplicated; the duplicate will be removed from the design.	No action needed unless you want to remove duplicate gene symbol entries from the input file.	The design can proceed without intervention.

ACCESSION NUMBER Errors

Message	Suggested action	Notes
The accession number is not found in the database.	The target has been removed. Dou-ble-check the accession number to ensure there are no typos in it.	The design can proceed without intervention, but the invalid target will be skipped.
The accession number is duplicated; the duplicate has been removed from the design.	No action needed, unless you want to remove duplicate accession num-ber entries from the input file.	The design can proceed without intervention.
The version of the accession is not found, but was replaced with the most recent version.	No action needed.	The design can proceed without intervention.
The transcript length is too short to generate a probe. The accession will be removed from the design.	Transcripts must be at least as long as one probe length in order to facili-tate a design. Shorter transcripts will not generate designs.	The design can proceed without intervention, but the invalid target will be skipped.

xGen Hyb Panel Design Tool

Technical support: applicationsupport@idtdna.com