Tagging, Enrichment & Next Generation Sequencing: Deep sequencing of specific regions of a large number of samples

Dr. Kerstin A. Stangier | Dipl.-Bioinf. Ulrike Schöck | Dr. Christopher Bauser
GATC Biotech AG, Jakob-Stadler-Platz 7, 78467 Konstanz, Germany, customerservice@gatc-biotech.com

Service workflow

Library preparation & Tagging

Starting material: genomic DNA

- Sample by sample enrichment → Mixing the samples for sequencing
- Mixing the tagged samples → Enrichment of mixed samples

Sequencing of mixture on Illumina Genome Analyzer $_{IIx}$

Bioinformatics:

Tag sorting
Quality trimming
Mapping
SNP analysis

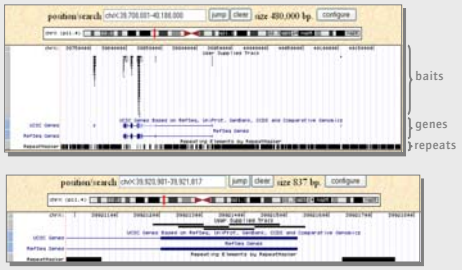
1. Library preparation from genomic DNA for Illumina Genome Analyzer$_{IIx}$, tagging

Ten samples were tagged using the proprietary method of GATC Biotech

Sample	Distribution in %
1	6.35
2	7.16
3	2.28
4	11.25
5	6.60
6	10.24
7	7.52
8	8.36
9	15.38
10	15.63
No Tag	9.22

Distribution of reads per tagged sample in percent

2. Bait/oligo design based on reference, repeat masking and quality check of baits



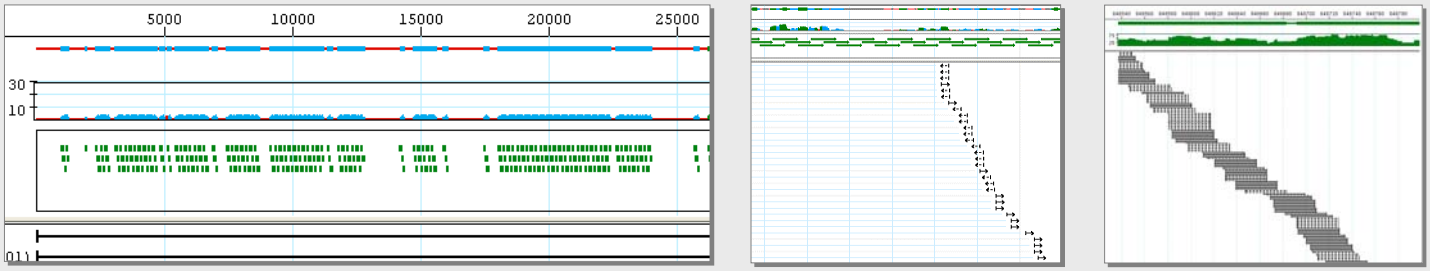
Quality check of baits in USCS browser

Quality check of baits in DNASTAR SeqBuilder (Lasergene)

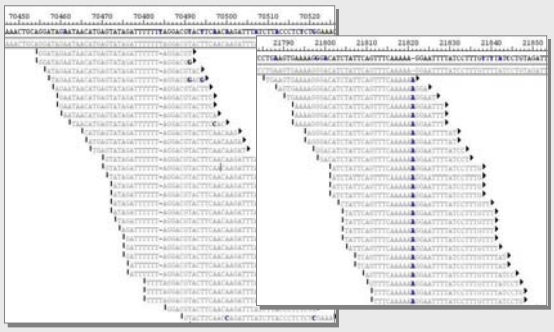
Region	Size / Mbp	%
genomic	3.2	100
repeat	1.7	51.2
enriched	1.5	48.8

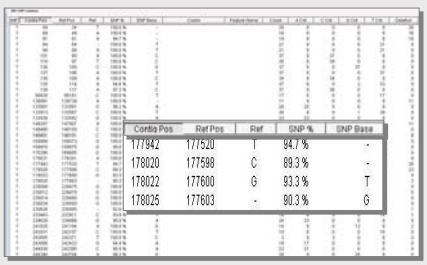
Percentage of repeats in targeted region of the human genome

3. Enrichment, sequencing and bioinformatic analysis



Mapped reads in DNASTAR SeqMan Pro (Lasergene)





SNP analysis in DNASTAR SeqMan Pro (Lasergene)

Position	Coverage	Reference	Consensus	Qual(G)	Qual(A)	Qual(T)	Qual(C)	Qual(N)
41649	19	C	T	0	0	313	257	9
41726	11	A	G	224	102	0	0	12
42339	21	A	C	0	189	0	352	35
43461	24	C	T	0	0	253	180	29
43896	22	C	T	0	0	345	273	24
45140	21	A	G	324	206	0	0	22
45369	11	T	G	156	0	80	0	11
45559	18	C	A	0	235	0	212	32
45636	24	C	T	0	0	266	260	18
45748	24	G	A	217	378	0	0	25
51426	15	A	G	201	181	0	0	9
72440	17	T	C	0	0	0	547	16
86634	4	G	C	21	0	0	26	8
96307	26	A	C	0	0	0	715	17
96350	6	T	C	0	0	0	153	0
96353	4	T	G	80	0	0	0	12
96393	6	A	C	0	0	0	169	7
96408	5	G	C	0	0	0	17	2
102433	11	T	G	98	0	80	0	21
113988	14	T	C	0	0	117	223	21
114681	25	C	T	0	0	484	322	10
151966	26	A	G	999	0	0	0	0

SNP detection table by GATC Biotech

Conclusion

1. Enrichment strategies can be used for the analysis of genomes and transcriptomes.
2. Tagging of samples allows for the combination of any number of samples with any size of targeted region to be sequenced.
3. Sample tagging/barcoding decreases sequencing costs.