Core facility for genotyping

Eivind Hovig[1,2,3]

Ambition:

- To provide all necessary services within the medical domain at a competitive price
- To develop and maintain competence in medical genotyping
- To insure rapid and simple access for relevant projects

Content

- Genotyping strategies
  - Platforms
  - Linkage (family analysis)
  - Association studies
  - Copy number variation (CNV)
- Bioinformatics
  - SNP information sources
  - Tools
- Core facility services
  - Organization
  - Funding
  - Pricing
  - Capacity/availability

Single nucleotide polymorphisms

Distribution of **5.3 mill. biallelic** validated SNPs in different regions of the human genome:

- *64.5% in intergenic regions*
- *33.7% in RefSeq introns*
- *0.7% in RefSeq untranslated regions (utr)*
- *1.1% in RefSeq exons*
  - *Synonymous (~0.55%)*
  - *Non-synonymous (~0.55%)*

Platforms

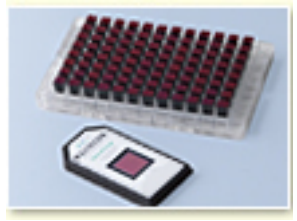
illumina®

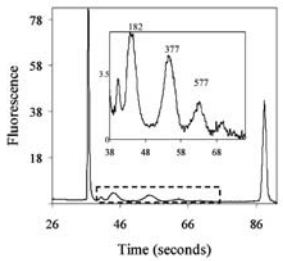
Infinium II assay provides a complete genomic solution with arrays containing from 300 000 to 1 million SNP. BeadChips come in Duo (2-sample) or Quad (4-sample) formats .

Goldegate provide assays with a focused content or custom panels. GoldenGate Panels are available in panels of 96 and 384–1,536 assays per OPA tube.

AFFYMETRIX®

can provide both **Whole Genome** arrays for linkage analysis, association studies, population genetics, chromosomal copy number and **Targeted Genotyping** for fine mapping, and custom SNPs.

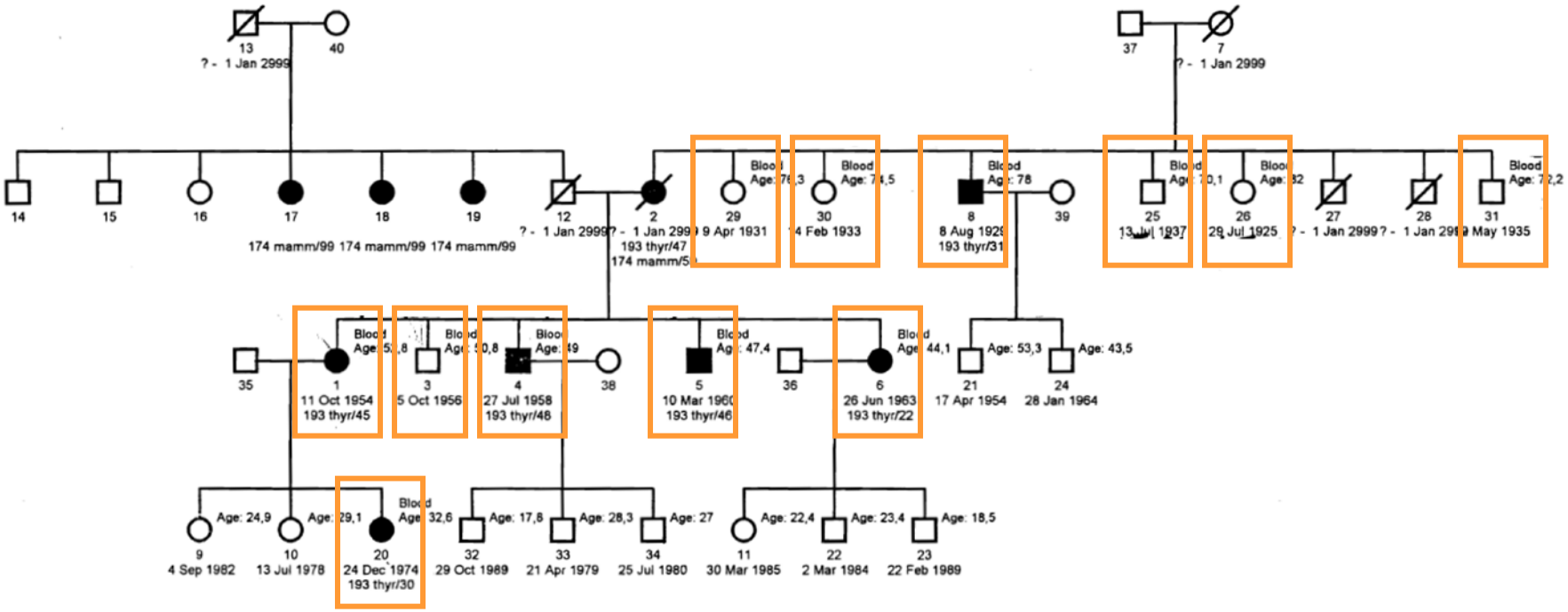




Cycling Temperature Capillary Electrophoresis is a melting gel technique modified by Per O. Ekstrøm for high-resolution capillary electrophoresis. The assay which cost effective, flexible and offers the possibility of genotyping as well as screening the target sequence for unknown DNA variants both with single samples or in bulk screening (500 samples).

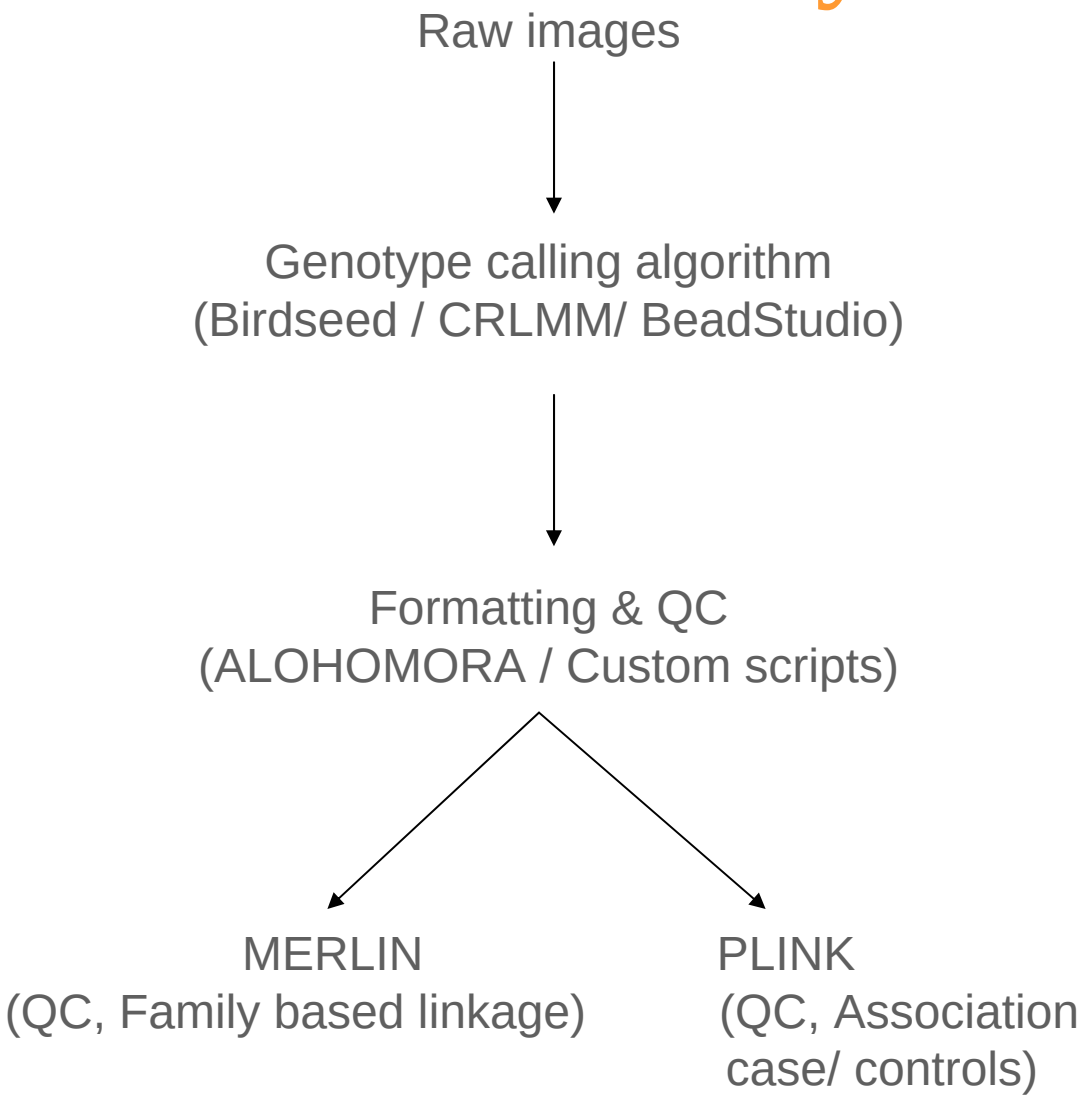
Example of use: Linkage analysis

Family with history of papillary thyroid carcinoma

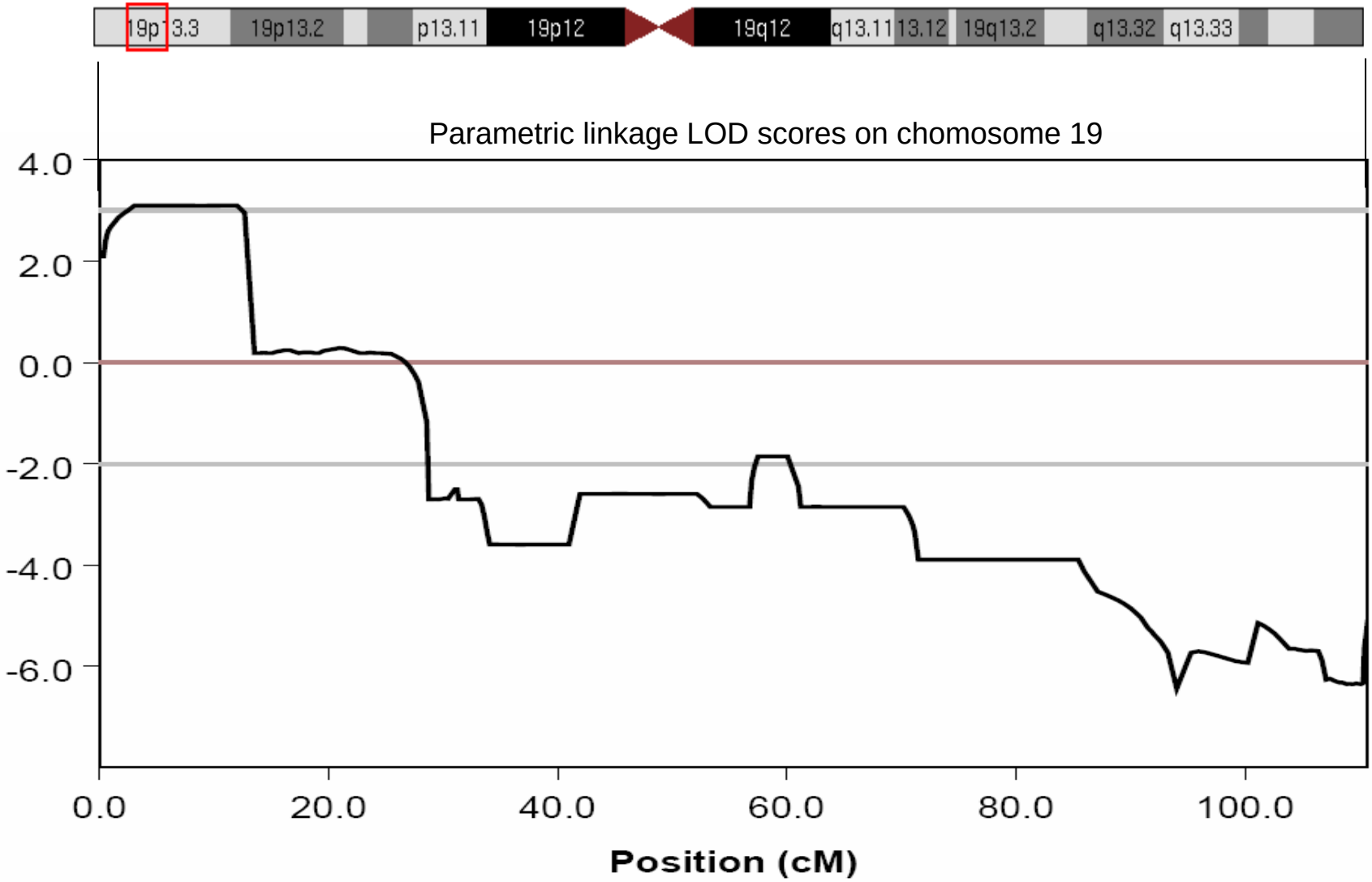


Collaboration with thyroid cancer group, and L. Mæhle, P. Møller

Pipeline for SNP arrays:

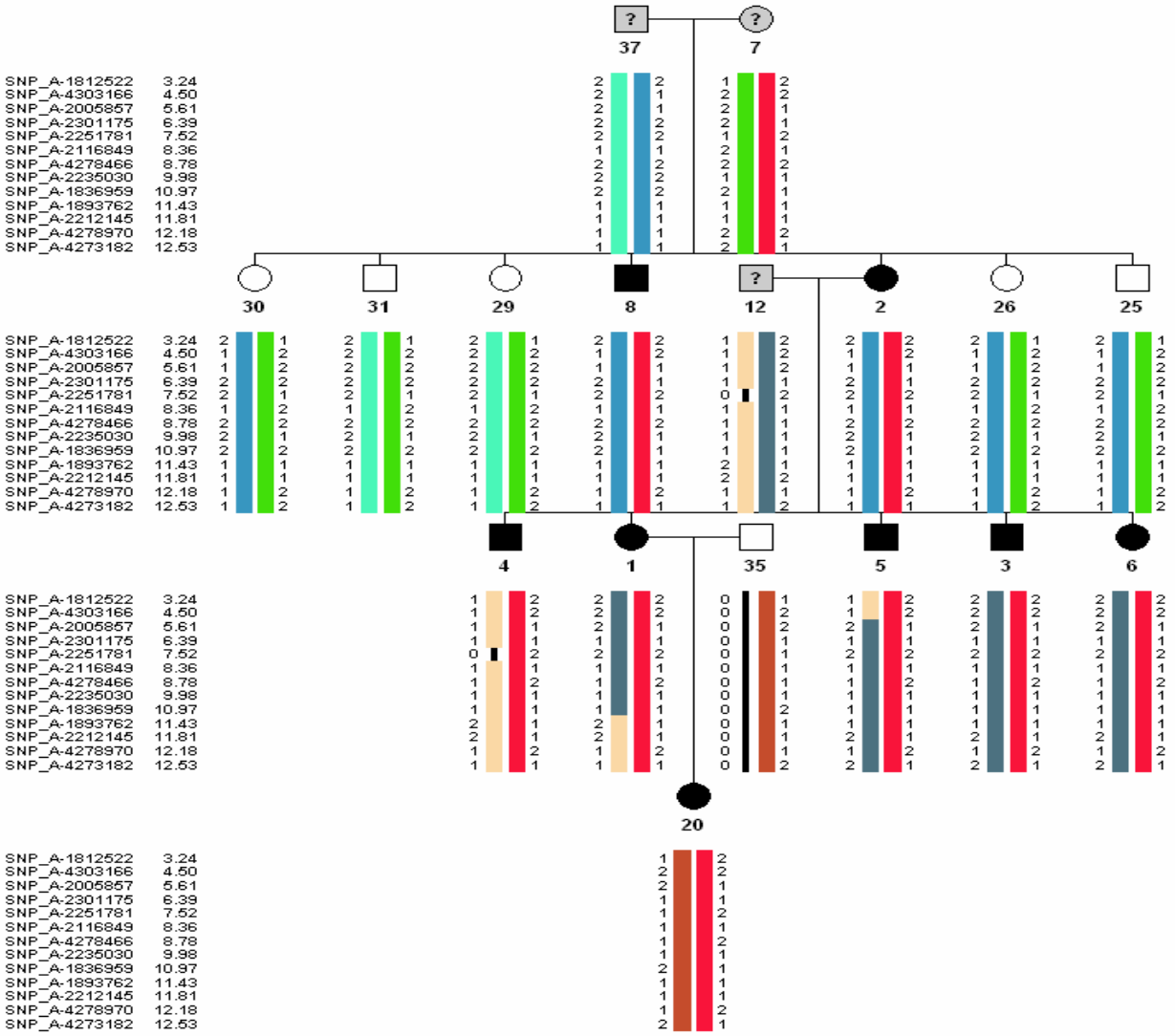


Region defined by recombination

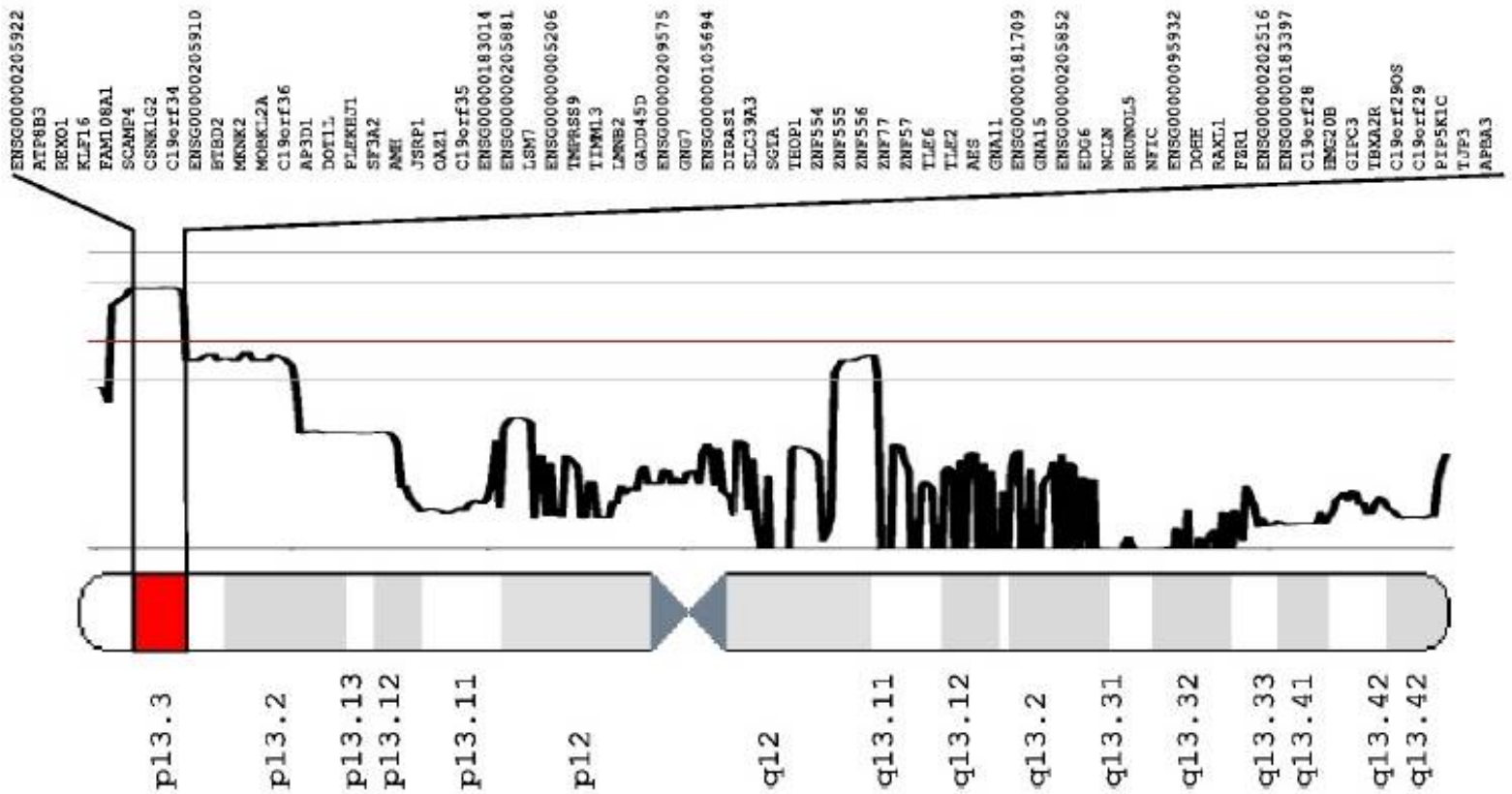


Family haplotypes

2.460.660 bases
included in region



Which gene?


ENSG00000205922
ATP6B3
REXO1
KLF16
FAM108A1
SCAMP4
CSNK1G2
C19orf34
ENSG00000205910
BTBD2
MRNK2
MOBKL2A
C19orf36
AP3D1
DOT1L
PLEKHJ1
SF3A2
AMH
JSRP1
GAZ1
C19orf35
ENSG00000183014
ENSG00000205881
LSM7
ENSG00000005206
TMPRSS9
TIMM13
LMNB2
GADD45D
ENSG00000209575
GNG7
ENSG00000105694
DIRAS1
SLC39A3
SGTA
TEDP1
ZNF554
ZNF555
ZNF556
ZNF77
ZNF57
TLE6
TLE2
AES
GNA11
ENSG00000181709
GNA15
ENSG00000205852
EDG6
NCLN
BRUNOL5
NFIC
ENSG00000095932
DOHH
RAKL1
FGR1
ENSG00000202516
ENSG00000183397
C19orf28
EMG20B
GIPC3
TBXA2R
C19orf290S
C19orf29
PIP5K1C
TJP3
APBA3
p13.3
p13.2
p13.13
p13.12
p13.11
p12
q12
q13.11
q13.12
q13.2
q13.31
q13.32
q13.33
q13.41
q13.42
q13.42

Priority genes:

Sequenced:
GADD45B - Growth arrest and DNA-damage-inducible protein GADD45 beta
HMG20B - SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1-related
DIRAS - GTP-binding protein Di-Ras1

Nothing identified so far...

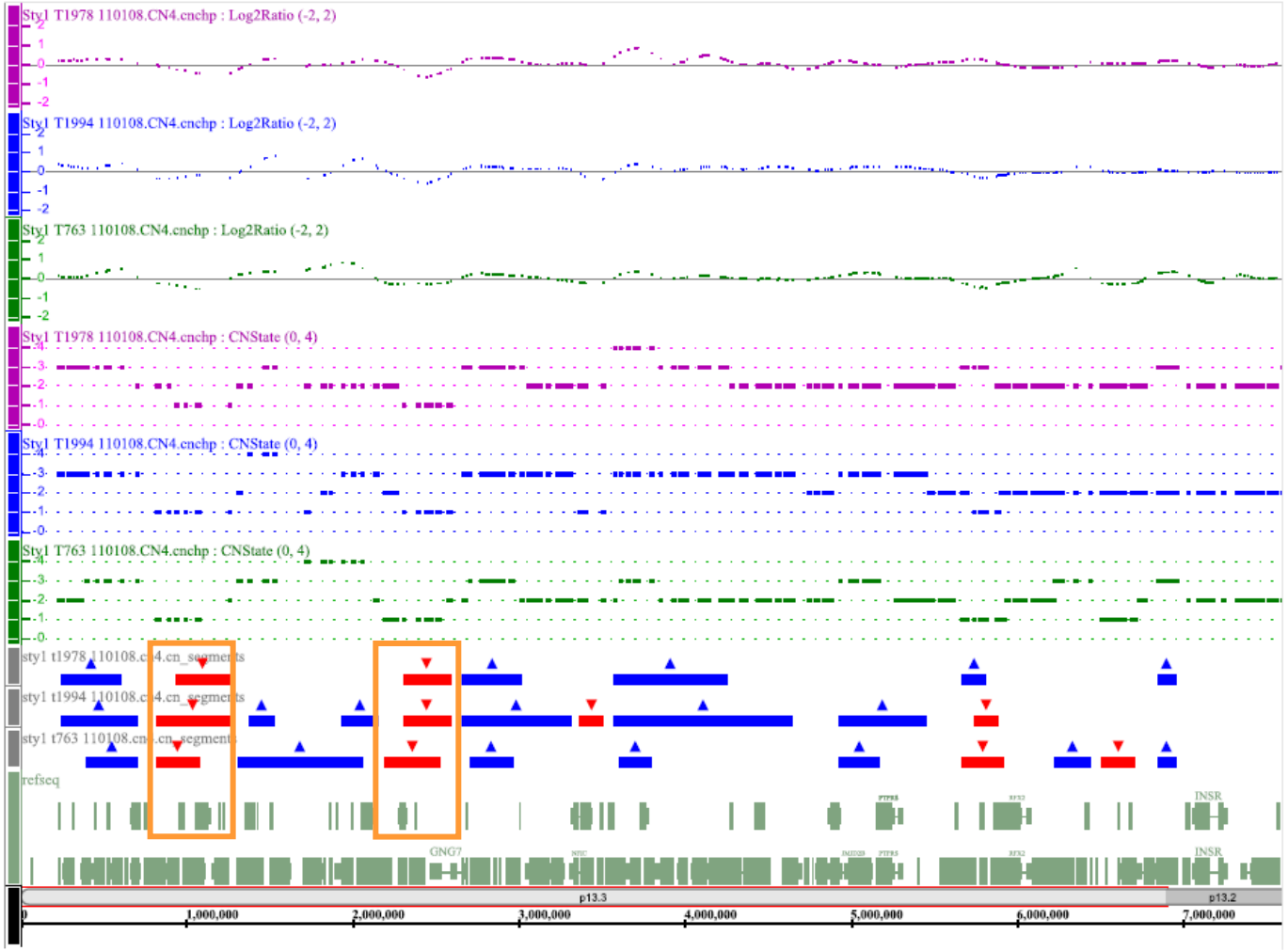
Other possible candidates:
MKNK2 - MAP kinase-interacting serine/threonine-protein kinase 2
GNG7 - Guanine nucleotide-binding protein gamma-7 subunit
GNG11 - Guanine nucleotide-binding protein subunit alpha-11
GNA15 - Guanine nucleotide-binding protein alpha-15 subunit

Help in CNV in tumors?

DNA extracted from paraffin sections and amplified

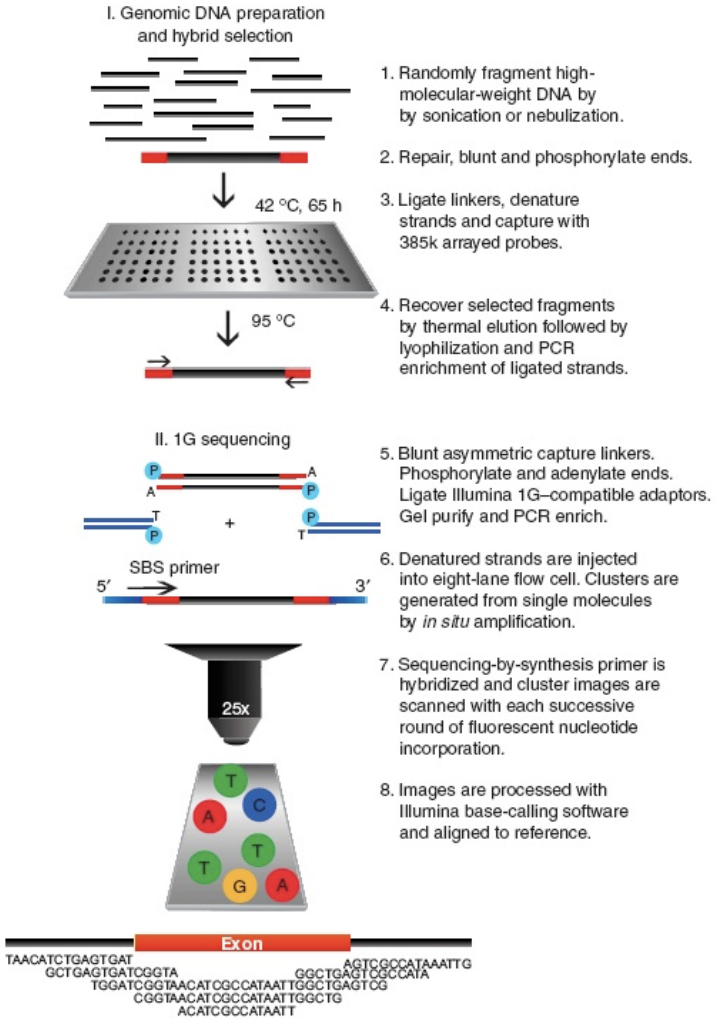
Thus, low call rate

Chromosome 19p13.3 CNV, 3 tumor vs 12 normal tissue



RIKSHOSPITALET

Massive sequencing option:



Hybrid capture opens for selecting
defined megabase regions for sequencing with high coverage

Hodges et al. Nature Genetics, 39(12) December 2007

Association analyses

- Comparison of disease versus control for overrepresented markers of disease
- Complex traits require large groups for statistical power (~1000 per group)
- Requires strong statistical and bioinformatics effort
- Statistical genomics consulting (Arnoldo Frigessi and associates) through the bioinformatics core facility

SNP bioinformatics research:

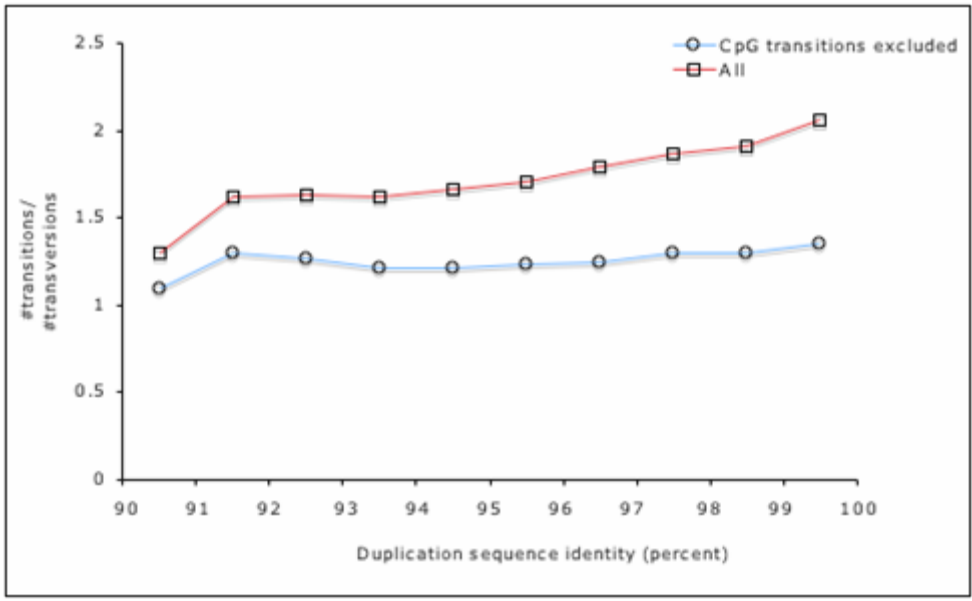
Example: Mutation information from segmental duplications

Detect mismatches (point mutations) in DNA duplicon sequence alignments

Exclude mutations in high-copy repeats (as annotated by RepeatMasker

Exclude mutations occurring in RefSeq transcripts

340,000 point mutations



Core facility service

Web: http://core.rr-research.no

Being organized in concert with the national resources through FUGE2, with Rikshospitalet coordinating human genotyping

Other members:

- Bergen NMC
- Trondheim NMC
- CIGENE, UMB, funded specifically with 10MNOK (responsible for production biology)
- CEES, UoO, funded for high throughput sequencing
- FUGE bioinformatics platform

Funding

Recently established and currently unfunded

Platforms through the Radium Hospital legacies and through support from the Oslo node of NMC and the Rikshospitalet microarray core facility

Thus, any project must by definition be financed by interested and willing users!

Rough pricing estimates for service:

Platform	Assay	Price per sample in NOK
Illumina Infinium II	Human1M-Duo v3.0 DNA Analysis Kit	3850,-
Certified	Human610-Quad v1.0 DNA Analysis Kit	2227,-
	HumanCNV370-Quad v3.0 DNA Analysis Kit	1657,-
Illumina Goldengate	Custom panel	312,-
Affymetrix	Genome-Wide Human SNP Array 6.0	2070,-
	Mapping 500K Array Set	2080,-
CTCE		1,- (per SNP)

RIKSHOSPITALET

Maximal present yearly capacity: 4300 Infinium chips.

Personnel

Genotyping:

Karen-Marie Heintz

Bioinformatics core:

Morten Johansen

Morten Mattingsdal

Vegard Nygaard

Halfdan Rydbeck

Torbjørn Rognes

Eivind Hovig

Research:

Sigve Nakken