

## Sequencing Pancreatic Cancer

The Australian Pancreatic Cancer Genome Initiative


Andrew Biankin




Garvan Institute of Medical Research, 384 Victoria Street Darlinghurst NSW 2010

[www.garvan.org.au](http://www.garvan.org.au)

## Australian ICGC



**ICGC Goal:**  
Generate comprehensive catalogues of genomic abnormalities in tumors (n=500) from 50 different cancer types (and/or subtypes) which are of clinical & societal importance across the globe

<http://www.icgc.org/>

**Sequencing + genome analysis**  
Sean Grimmond  
QCMG, Brisbane

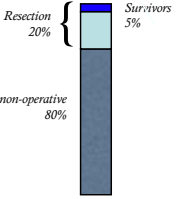
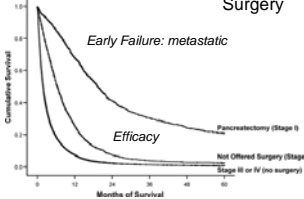
**370 Pancreatic cancers**  
Andrew Biankin  
Garvan Institute, Sydney

**130 Ovarian cancers**  
David Bowtell  
Peter Mac Cancer Centre, Melbourne

## Pancreatic Cancer



- 4<sup>th</sup> cause of cancer death (5% alive at 5 years)
- 90% die within a year
- Mortality unchanged for almost 50 years
- Usually advanced at diagnosis
- Surgery only potential for cure – no ability to predict those that will benefit
- Chemotherapy modestly effective – but some efficacy in undefined subgroups

## Treatment


- Chemotherapy: may work in undefined subgroups

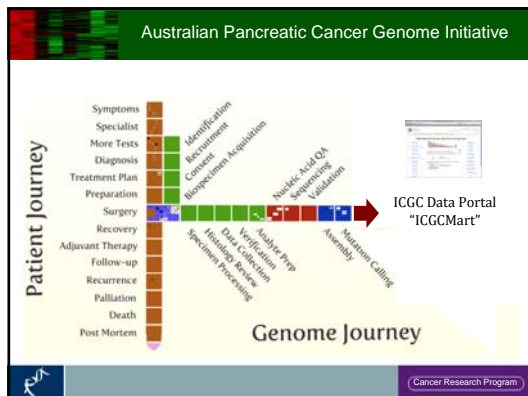
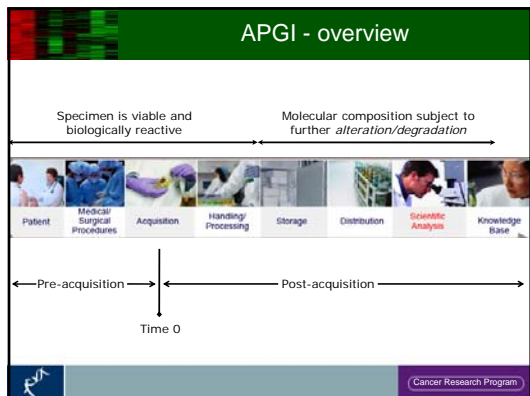
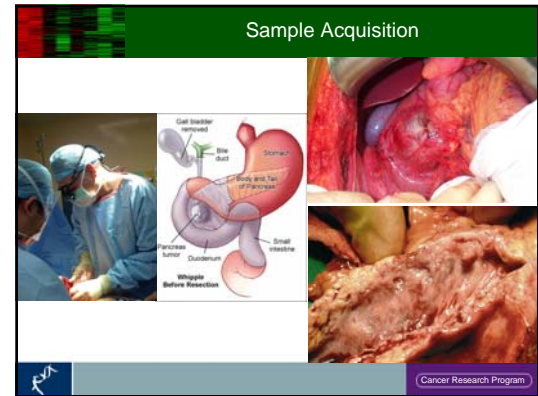
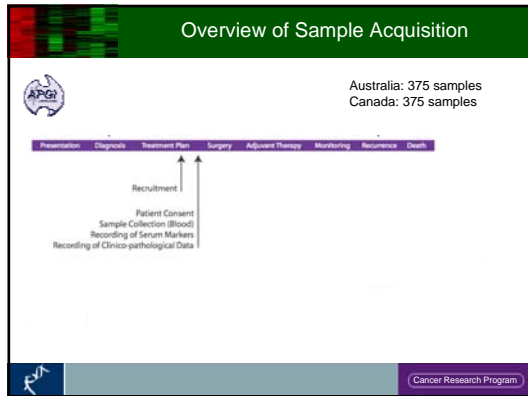
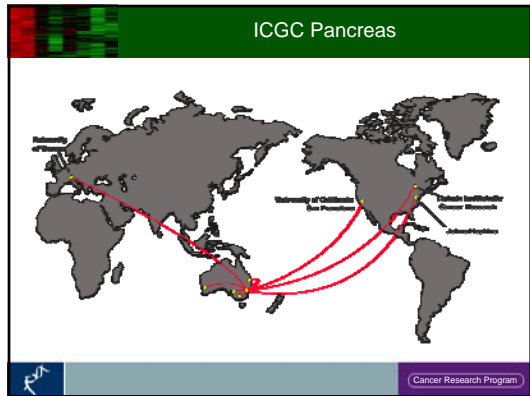
## Australian Pancreatic Cancer Genome Initiative

- ICGC Australian Pancreatic Cancer Genome Initiative – Aim is to fully characterise the genomic, epigenomic and transcriptomic aberrations in a large number of pancreatic cancer
- Relate to clinico-pathological and treatment parameters
- Generate primary xenografts and cell lines for functional assessment, therapeutic and companion biomarker preclinical testing.

## Australian Pancreatic Cancer Genome Initiative





### Queensland Centre for Medical Genomics

#### SOLiD Instrument Room

800m<sup>2</sup> refurbished labs: Dedicated labs for workflow (library prep, PCR, post PCR, validation lab). Environmentally controlled sequencing, array facility, LCM core (July 09)

11 SOLiD sequencers: Have been early adopters of SOLiD sequencing. Includes capacity for R&D as well as HTS sample processing (Aug-Nov 09).

HPC & Storage: Dedicated HPC (in partnership with SGI) 200 node cluster, 1.2Tb RAM, 100Tb scratch space, 1 Pb of storage )

UQ HPC

Cancer Research Program

## Sequencing Strategy

| Genome  | Exome   | Transcriptome   | Epigenome   |
|---|---|---|---|
| <p>SNP/CNV chip analysis, gDNA sequencing</p>       | <p>Agilent Sure select, exome sequencing</p>            | <p>Expression array, miRNAseq, miRNAseq</p>                           | <p>Methyl Miner enrichment, methyl seq</p>                    |
| <p>Tumor tissue &amp; normal<br/>30 fold (75Gb)</p> | <p>Tumor tissue &amp; normal<br/>&gt;100 fold (5Gb)</p> | <p>Tumor tissue &amp; adjacent normal<br/>~100million reads (5Gb)</p> | <p>Tumor &amp; adjacent normal<br/>~20million reads (1Gb)</p> |

## Genomic aberrations in cancer

- Single Nucleotide variations (SNV)
- Small insertions and deletions (INDELS)
- Copy number changes
- Large Chromosome rearrangements
- Epigenome

**Simple Nucleotide Variations**  
nonsense mutations, missense mutations, frameshift mutations

**Insertions**  
copy number variations, amplifications of oncogenes

**Translocations**  
fusion genes, aberrant gene expression, gene dysregulation

**Deletions**  
copy number variations, deletions of tumour suppressors

**Epigenomic**  
eg. methylation, histone modification

## Single Nucleotide Variants (SNV's)

Map tags to genome (Bioscope)

Identification of SNV (diBayes)

Annotate SNVs (eg. dbSNP, non-synonymous somatic)

Rank SNVs (eg. Polyphen, ConPredict)

Validate SNVs (eg. SNP chip, Sanger Sequencing)

**Somatic or germline variant**

**Comparison to dbSNP 130**

**SNV consequence:**

e.g. if in an ORF non-synonymous (V234K, 1234T>A) splice site 5'/3'UTR stop gained/lost

local Ensembl DB install

**Pfam domain annotation**

## Identification of SNVs

Map tags to genome

Identification of SNV

Annotate SNVs (eg. dbSNP, non-synonymous somatic)

Rank SNVs (eg. polyphen, ConPredict)

Validate SNVs (eg. SNP chip, Sanger Sequencing)

**All potential SNVs are visualized in IGV**

**Tumour gDNA** C/T

**Normal gDNA** C/C

## Transcriptome informing SNV analysis

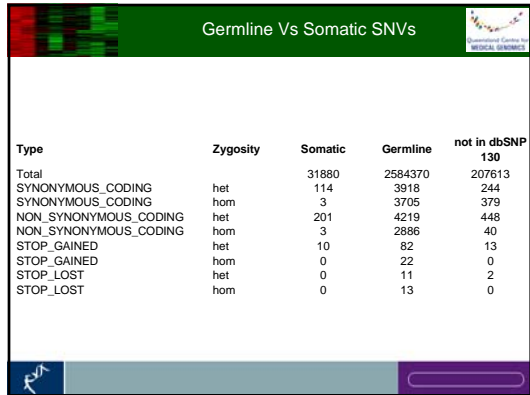
**Tumour DNA: A/G**  
Non-synonymous (N456S)  
CTNBP2HL  
Polyphen: probably damaging - deleterious

**Normal DNA: A/A**

**Variant expressed**  
Tumour DNA: A and G expressed

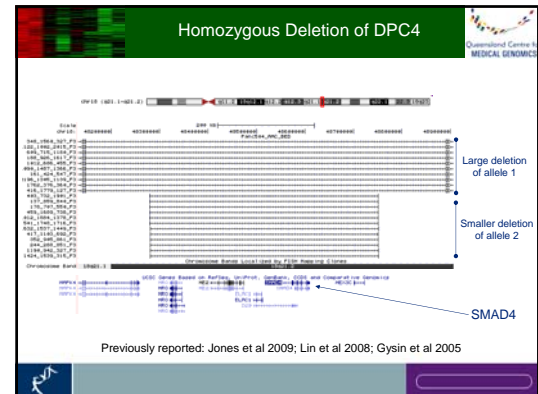
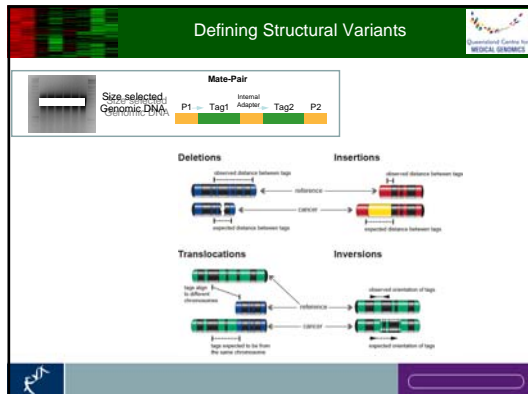
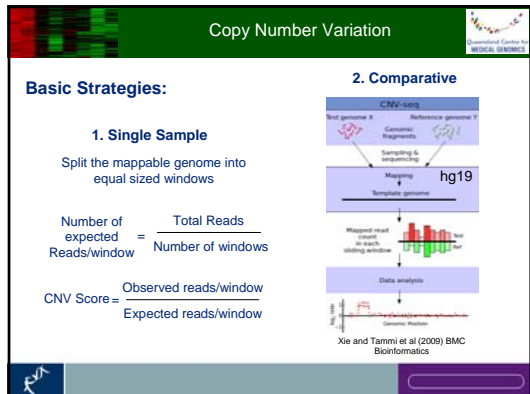
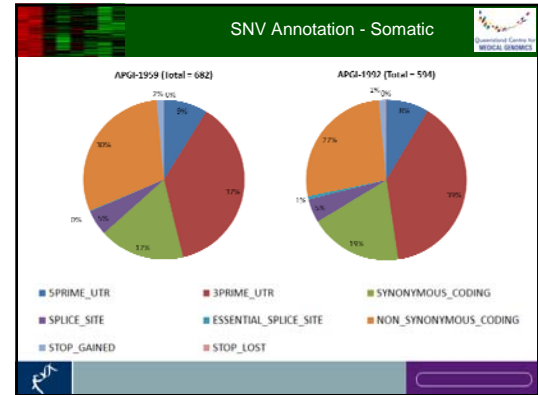
## Sequences to date

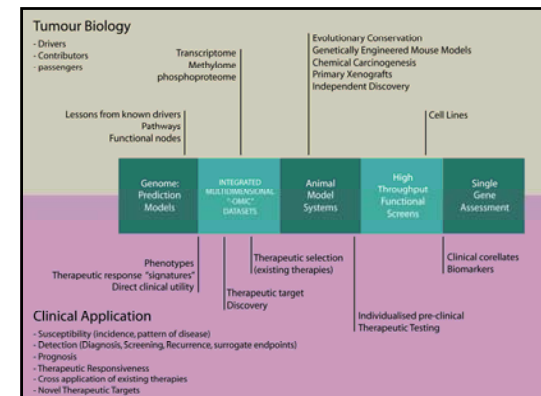
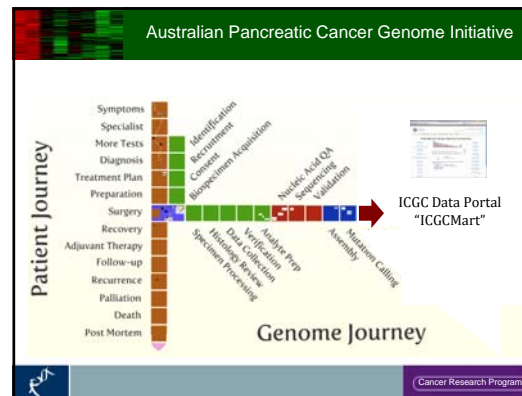
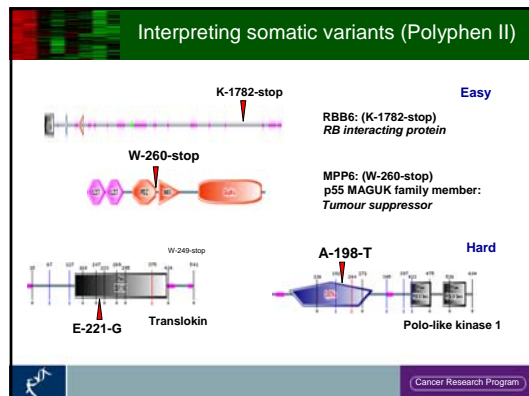
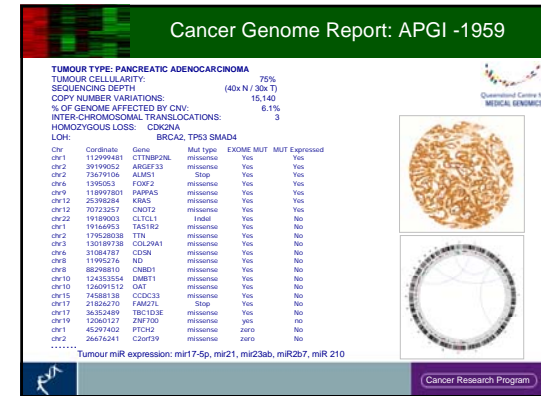
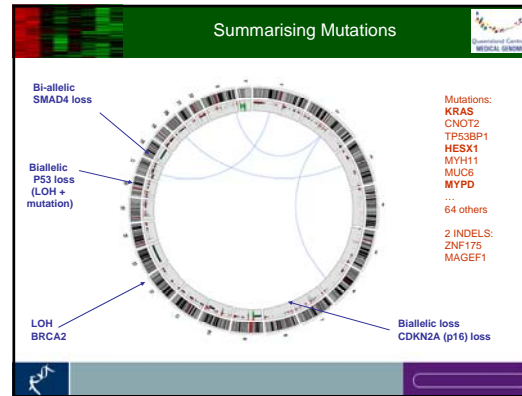
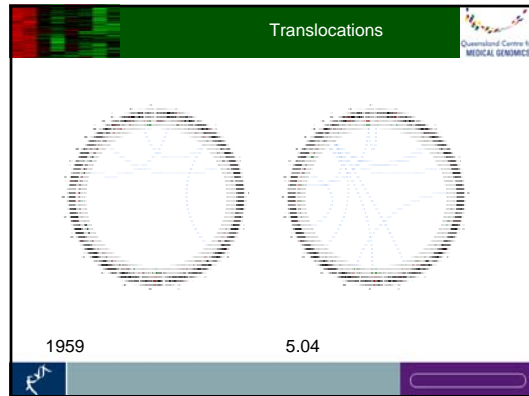
|                 |   |
|-----------------|---|
| Whole Genome:   | 7 Pancreatic adenocarcinomas<br>3 Pancreatic cell lines<br>10 matched normal samples: |
| Exome:          | 10 Pancreatic adenocarcinomas<br>2 pancreatic cell lines<br>12 matched normal samples |
| Transcriptomes: | 12 pancreatic tumours<br>3 pancreatic cell lines<br>3 normal adjacent pancreas        |



### SNV Annotation - Somatic

| Type                   | APGI-1959 | APGI-1992 |
|------------------------|-----------|-----------|
| INTERGENIC             | 16292     | 17158     |
| DOWNSTREAM             | 3517      | 3668      |
| UPSTREAM               | 3429      | 3518      |
| INTRONIC               | 10399     | 9755      |
| WITHIN_NON_CODING_GENE | 1633      | 2042      |
| 5PRIME_UTR             | 60        | 51        |
| 3PRIME_UTR             | 255       | 232       |
| SYNONYMOUS_CODING      | 117       | 111       |
| SPLICE_SITE            | 35        | 28        |
| ESSENTIAL_SPLICE_SITE  | 1         | 4         |
| NON_SYNONYMOUS_CODING  | 204       | 159       |
| STOP_GAINED            | 10        | 8         |
| STOP_LOST              | 0         | 1         |
| TOTAL                  | 31880     | 32715     |





## Opportunities

- Drugs with known target mechanisms/signatures
  - Preclinical testing
  - Direct clinical application
- Therapeutic Target Discovery
  - Resistance Mechanisms
  - Novel Mechanisms
- Recurrence Detection
  - Surrogate Endpoints for Clinical Trials
- Companion Biomarker Discovery and Validation
  - Preclinical models

Cancer Research Program

## Molecular Guided Therapy for Pancreatic Cancer

*Personalised Medicine*

*The right drug, for the right patient, the first time*

*Using a drug that might work when all others have failed*

Cancer Research Program

## Molecular Phenotype Guided Therapy

**Biomarker (definition)**

"A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathological processes, exposure to environmental and lifestyle factors, or responses (pharmacological or otherwise) to a therapeutic intervention"

- Select best therapy
- Focus clinical trials
- Investigate resistance
- Better inform our patients

**Personalised Therapy**

Cancer Diagnosis

Biomarker/Signature

Responds to treatment 'X'

1

25% response

2

18% response

3<sup>x</sup>

12% response

4

6% response

→ Toxicity, delay, cost

Cancer Research Program

## Molecular Phenotype Guided Therapy

Cancer Research Program

## Individualised Molecular Pancreatic Cancer Therapy (IMPACT)

Patient Journey

- Symptoms
- Specialist
- More Tests
- Diagnosis
- Treatment Plan
- Preparation
- Surgery
- Recovery
- Adjuvant Therapy
- Follow-up
- Recurrence
- Palliation
- Death
- Post Mortem

Genome Journey

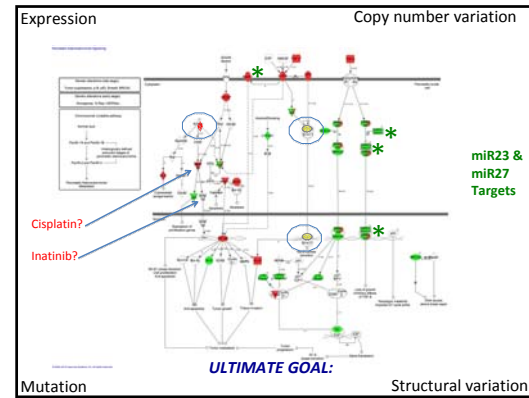
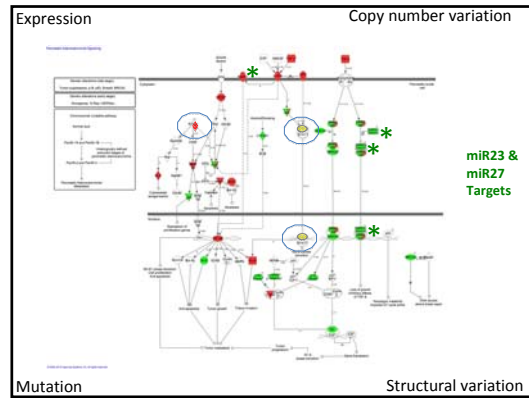
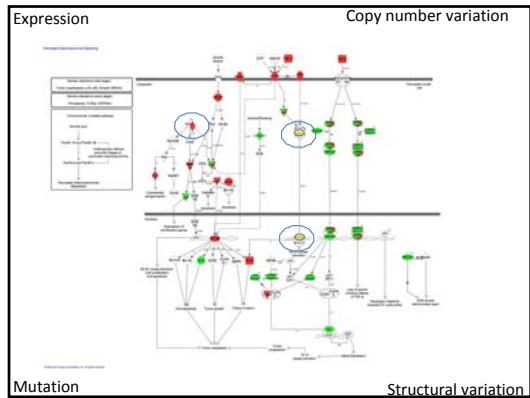
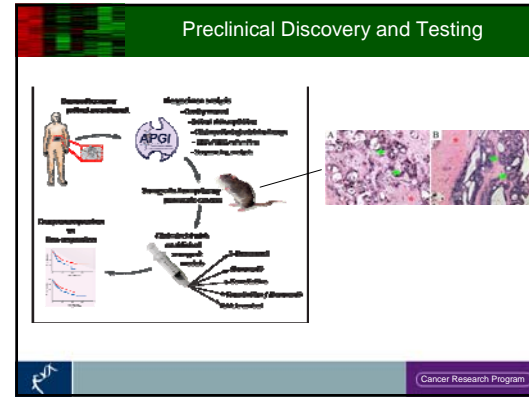
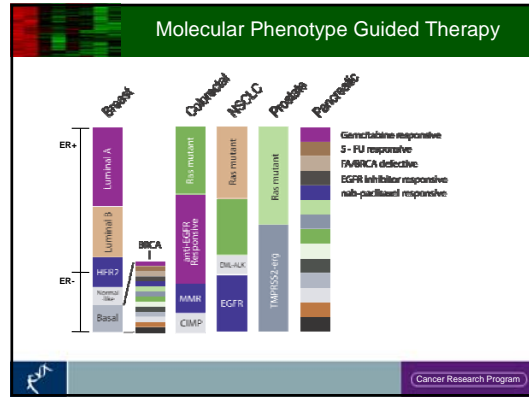
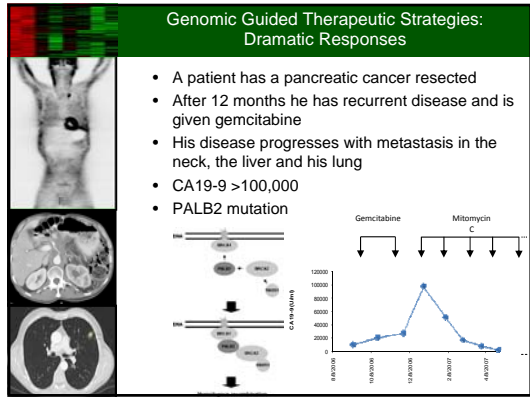
- Identification
- Recruitment
- Consent
- Biospecimen Acquisition
- Nucleic Acid QA
- Sequencing
- Validation
- Assembly
- Functional Insights
- Clinical Considerations
- Clinical Validation
- Translation
- Implementation
- Clinical Testing

Inform Chemotherapy

Cancer Research Program

## Overview of Sample Acquisition

Cancer Research Program



**IMB Institute for Molecular Bioscience**

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

SYDNEY SOUTH WEST ACQA HEALTH SERVICE NSW HEALTH

Queensland Centre for MEDICAL GENOMICS

GARVAN INSTITUTE

APGI

JOHNS HOPKINS MEDICINE

Cancer Council NSW

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Queensland Government

AB applied biosystems | sgi

UCSF

NSWPCN

UNIVERSITY OF SYDNEY

### Acknowledgements

**OCMG**

**Sean Grimmond**

Peter Wilson

**Genome Biology:**  
Karin Kasahn  
Alic Waddell  
Nicola Cloonan  
Shivangi Wani  
Nick Walling  
Koorilana Krishnan  
Malissa Brown  
Mathi Thiagarajan  
Illi Shepherd  
Mary Martin

**Bioinformatics:**  
John Pearson  
Liam Taylor  
David Tang  
Craig Nourse  
Suzanne Manning  
Shsan Nourbakhsh  
Liam Hartigan  
Matt Anderson  
David Wu  
Scott Wood  
William Waterford  
Dile Holme

**Genome Sequencing:**  
Brooke Gardiner  
David Miller  
Craig Nourse  
Suzanne Manning  
Shsan Nourbakhsh  
Liam Hartigan  
Senei Idrisovic

**Life Technologies:**  
Gabriel Kolte  
John Davis  
Kevin McKernan  
Clarence Lee  
Jian Gu  
Eileen Dimallata

**HPV (QO):**  
Lutz Pross  
Ziping Fang  
David Green  
Chris Toon

**Silicon Graphics:**  
Nick Gornoni  
Todd Churchwood  
Gerald Hafer

**UC Microarray Facility:**  
Kiana Nohak  
Rebecca Foale

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Cancer Research Program

Queensland Government

NSW Health and Medical Research Council

UNIVERSITY OF SYDNEY

**GARVAN INSTITUTE**

Rob Sutherland  
Sue Henshall  
Liz Musgrove  
Roger Daly  
James Kench

Amber Johns  
Michelle Thomas  
Chris Toon  
Mary-Anne Brancato  
Cathy Axford

Chris Scarlett  
Marc Jones  
David Chang  
Jianmin Wu  
Anthony Gill  
Page Tobelman  
Jeremy Humphris

**JOHNS HOPKINS**

Ralph Hruban  
Jim Eshleman

Anirban Maitra  
Chris Iacobuzio-Donahue

**VERONA**

Aldo Scarpa  
Claudio Bassi  
Paolo Pedersoli

**UCSF**

Margaret Tempero  
Rita Lawlor

Emily Colvin  
Amanda Mawson  
Johana Susanto  
Marina Pajc  
Mona Martyn-Smith  
Lorraine Chantrill  
Adnan Nagrial  
Venessa Chin

... and MANY, MANY MORE ...

**APGI**  
Australian Pancreatic  
Cancer Genome Initiative

Garvan Institute of Medical Research, 384 Victoria Street Darlinghurst NSW 2010

www.garvan.org.au

### ICGC

**ALL TOGETHER NOW**  
Eleven countries have signed on to sequence DNA from 500 tumour samples for each of more than 20 cancer types for the International Cancer Genome Consortium. Each cancer type is estimated to cost nearly US\$20 million to sequence.

**Number of cancer types being sequenced:**

- Canada: 1 (Pancreatic cancer (ductal adenocarcinoma))
- United States: 6+ (Through the Cancer Genome Atlas: Ovarian cancer, Brain cancer (glioblastoma multiforme), Lung cancer (adenocarcinoma and squamous), Liver cancer (hepatocellular carcinoma), Acute myeloid leukaemia, Colon cancer (adenocarcinoma), China)
- Spain: 3 (Chondrosarcoma, Squamous cell carcinoma, Leukemia)
- France: 3 (Breast cancer (HER2 overexpressing), Liver cancer (hepatocellular carcinoma), Renal cell carcinoma (adenocarcinoma))
- Germany: 1 (Pancreatic cancer (ductal adenocarcinoma), Breast cancer (ER+, PR+, HER2-), Breast cancer (ER-, HER2-), European Liver neoplasms)
- India: 1 (Ovarian cancer (epithelial))
- Italy: 1 (Pancreatic cancer (adenocarcinoma), Endometrial cancer, Cervical cancer)
- China: 1 (Pancreatic cancer (ductal adenocarcinoma))
- Japan: 1 (Liver cancer (hepatocellular carcinoma))
- Australia: 2 (Pancreatic cancer (ductal adenocarcinoma), Ovarian cancer)

### Surrogate Endpoints in Clinical Trials

Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing  
Petrova, J. et al. *SG Trials* Med Z, 2014:14 (2014)

Percentage of personalized DNA in plasma

Time (days)

0 10 11 14 35 137

Sigmoid colectomy

Chemotherapy

Right hepatectomy

0% 20% 40% 60%

Control plasma

CH-1: 1 personalized marker  
CH-2: 1 personalized marker

| Sample Type    | CH-1 | CH-2 |
|----------------|------|------|
| Patient tumor  | 100% | 100% |
| Patient normal | 0%   | 0%   |
| Patient plasma | ~30% | ~30% |
| Control plasma | 0%   | 0%   |

## Analysis pipelines:

### Mapping to the human reference genome:

Bioscope for DNA  
Bioscope/X-Mate (Cloonan *et al* 2009 *Bioinformatics*) for mRNA  
miR-Mate for miRNA (N. Cloonan, unpublished)

### SNVs and small indels:

diBayes/small INDEL tool (Bioscope) for detecting SNVs/ indels  
In-house pipeline for calling somatic/germline events  
In-house pipeline for annotating events

### Copy number variants:

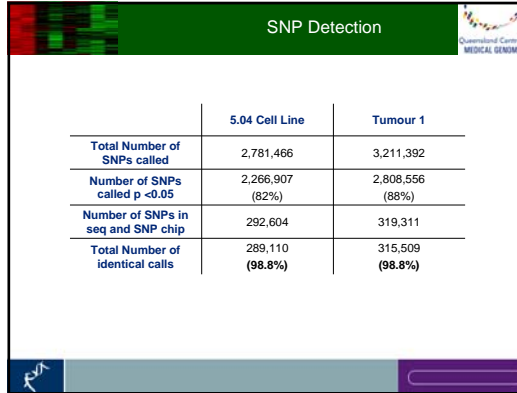
CNVSeq (Xie *et al* 2009 *BMC Bioinformatics*)  
In-house pipeline for calling somatic/germline events

### Structural variants:

In-house pipeline using long-mate pair data

### mRNA expression:

Bioscope WTA pipeline for RPKM values  
In-house pipeline for inferring transcriptional complexity



SNP Detection

|                                    | 5.04 Cell Line  | Tumour 1        |
|------------------------------------|-----------------|-----------------|
| Total Number of SNPs called        | 2,781,466       | 3,211,392       |
| Number of SNPs called $p < 0.05$   | 2,266,907 (82%) | 2,808,556 (88%) |
| Number of SNPs in seq and SNP chip | 292,604         | 319,311         |
| Total Number of identical calls    | 289,110 (98.8%) | 315,509 (98.8%) |

Quantitative Centre for MEDICAL GENOMICS