Update on The Cancer Genome Atlas for the National Cancer Advisory Board

Mark Guyer

Director, Division of Extramural Research, NHGRI

on behalf of the TCGA Project Team

September 7, 2010

The TCGA Project Team

THE CANCER GENOME ATLAS

- **NCI**
  - Ann Barker
  - Joe Vockley
  - Kenna Shaw
  - Laura Dillon
  - Greg Eley
  - Carl Schaeffer
  - Martin Ferguson
  - Peter Fielding
- **NHGRI**
  - Brad Ozenberger
  - Jacqueline Palchik
  - Jane Peterson
  - Peter Good
  - Elizabeth Thomson
  - Julia Zhang

The Directors: Harold Varmus, Eric Green, Francis Collins

Outline

- Review the last presentation to NCAB (Sept 2009)
- Summarize overall organization of TCGA and the tumors being targeted (plus the ICGC targets)
- Update on status of sample accrual and sequencing
- Future sample accrual goals
- Challenges
- Approaches to challenges

Background

THE CANCER GENOME ATLAS

- **Cancer is a disease of genomic alterations – identification of all genomic changes would enable defining cancer subtypes and generate a comprehensive set of alterations that characterize each type and subtype of cancer – potential to transform cancer drug discovery, diagnostics and prevention**

Background

THE CANCER GENOME ATLAS

- Cancer is a disease of genomic alterations – identification of all genomic changes would enable defining cancer subtypes and generate a comprehensive set of alterations that characterize each type and subtype of cancer – potential to transform cancer drug discovery, diagnostics and prevention

- **Following several workshops and a specific recommendation by the National Cancer Advisory Board, TCGA was launched as a collaboration between the NCI and NHGRI in 2006.**

Background

THE CANCER GENOME ATLAS

- Cancer is a disease of genomic alterations – identification of all genomic changes would enable defining cancer subtypes and generate a comprehensive set of alterations that characterize each type and subtype of cancer – potential to transform cancer drug discovery, diagnostics and prevention

- Following several workshops and a specific recommendation by the National Cancer Advisory Board, TCGA was launched as a collaboration between the NCI and NHGRI in 2006.

- **TCGA was initiated as a pilot designed to explore the processes needed to perform high-throughput, large scale disease-focused genome characterization, data integration and analysis**
  - **Biospecimens**
  - **Large-scale genome characterization and sequencing**
  - **Integration of data, laboratories and teams**
  - **Policies (e.g. data standards, data access, informed consent**

TCGA Pilot Program Conclusions

THE CANCER GENOME ATLAS

- **High-throughput large-scale comprehensive characterization is feasible and produces useful data**

TCGA Pilot Program Conclusions

- High-throughput large-scale comprehensive characterization is feasible and produces useful data
- **New cancer genes can be discovered**

TCGA Pilot Program Conclusions

THE CANCER GENOME ATLAS

- High-throughput large-scale comprehensive characterization is feasible and produces useful data
- New cancer genes can be discovered
- **Tumor subtypes can be differentiated based on comprehensive knowledge of genomic alterations**

TCGA Pilot Program Conclusions

- High-throughput large-scale comprehensive characterization is feasible and produces useful data
- New cancer genes can be discovered
- Tumor subtypes can be differentiated based on comprehensive knowledge of genomic alterations
- **Integrated teams can be built – and it will take teams to analyze multi-dimensional data**

TCGA Pilot Program Conclusions

- High-throughput large-scale comprehensive characterization is feasible and produces useful data
- New cancer genes can be discovered
- Tumor subtypes can be differentiated based on comprehensive knowledge of genomic alterations
- Integrated teams can be built – and it will take teams to analyze multi-dimensional data
- **This comprehensive approach can produce clinically relevant data**

TCGA Pilot Program Conclusions

- High-throughput large-scale comprehensive characterization is feasible and produces useful data
- New cancer genes can be discovered
- Tumor subtypes can be differentiated based on comprehensive knowledge of genomic alterations
- Integrated teams can be built – and it will take teams to analyze multi-dimensional data
- This comprehensive approach can produce clinically relevant data
- **At the level of individual genes, cancer genomics is complex. At the level of pathways, more coherence can be observed.**

TCGA Phase II: Overview

THE CANCER GENOME ATLAS

- **ARRA funds will be employed for 2 years to collect tissues for years 1-5 of TCGA – and scale up the Biospecimen Core Resource capacity**

TCGA Phase II: Overview

THE CANCER GENOME ATLAS

- ARRA funds will be employed for 2 years to collect tissues for years 1-5 of TCGA – and scale up the Biospecimen Core Resource capacity
- **During the two years of ARRA funding – plan to complete comprehensive genome characterization of 10 tumor types (at 200 cases/tumor type as a discovery set and more depending on tumor type; 180 exomes; 20 whole genomes/tumor), and partial characterization of 10 more (at 100 cases/tumor type)**
  - **GCCs will perform expression, copy number, methylation and miRNA characterization**
  - **Genome Sequencing Centers will use Nex-Gen sequencing technologies – exomes and whole genomes (cost dependent)**

TCGA Phase II: Overview

THE CANCER GENOME ATLAS

- ARRA funds will be employed for 2 years to collect tissues for years 1-5 of TCGA – and scale up the Biospecimen Core Resource capacity
- During the two years of ARRA funding – plan to complete comprehensive genome characterization of 10 tumor types (at 200 cases/tumor type as a discovery set and more depending on tumor type; 180 exomes; 20 whole genomes/tumor), and partial characterization of 10 more (at 100 cases/tumor)
  - GCCs will perform expression, copy number, methylation and miRNA characterization
  - Genome Sequencing Centers will use Nex-Gen sequencing technologies – exomes and whole genomes (cost dependent)
- **Genome Data Analysis Centers will integrate data from GCCs – GDAC-Bs will further integrate data, create new models and tools to refine and further add value to data for communities**

TCGA Project Pipeline

THE CANCER GENOME ATLAS

Supplementary Figure 1

Pathology QC

DNA & RNA Isolation, QC

Analysis

Sequencing

Expression, CNA & LOH, Epigenetics

Analysis

Data and Results Storage & QC

GDAC

Integrative Analysis

Comprehensive **Knowledge** of a Cancer Genome

= Process

= Data

= Results

= BCR = GSCs = CGCCs = DCC = GDACs

TCGA Project Pipeline

THE CANCER GENOME ATLAS

Supplementary Figure 1

Pathology QC

DNA & RNA Isolation, QC

Analysis

Sequencing

Expression, CNA & LOH, Epigenetics

Analysis

Data and Results Storage & QC

GDAC

Integrative Analysis

Comprehensive **Characterization** of a Cancer Genome

= Process

= Data

= Results

= BCR = GSCs = CGCCs = DCC = GDACs

TCGA Tumor Projects

THE CANCER GENOME ATLAS

Projects with comprehensive data available

- **Glioblastoma†**
- **Ovarian**

Projects in progress (partial data sets available)

- **Acute Myeloid Leukemia**
- **Colon Adenocarcinoma**
- **Rectal Carcinoma**
- **Lung Adenocarcinoma**
- **Lung Squamous Cell Carcinoma**
- **Renal Clear Cell Carcinoma**

TCGA Tumor Projects

THE CANCER GENOME ATLAS

Projects recently begun or upcoming

- Breast (multiple types)*
- Bladder
- Cervical
- Head and neck
- Liver
- Lymphoma
- Melanoma
- Multiple myeloma
- Pancreatic
- Prostate
- Sarcoma
- Stomach*
- Thyroid
- Uterine* (endometrial)

ICGC Projects (June, 2010)

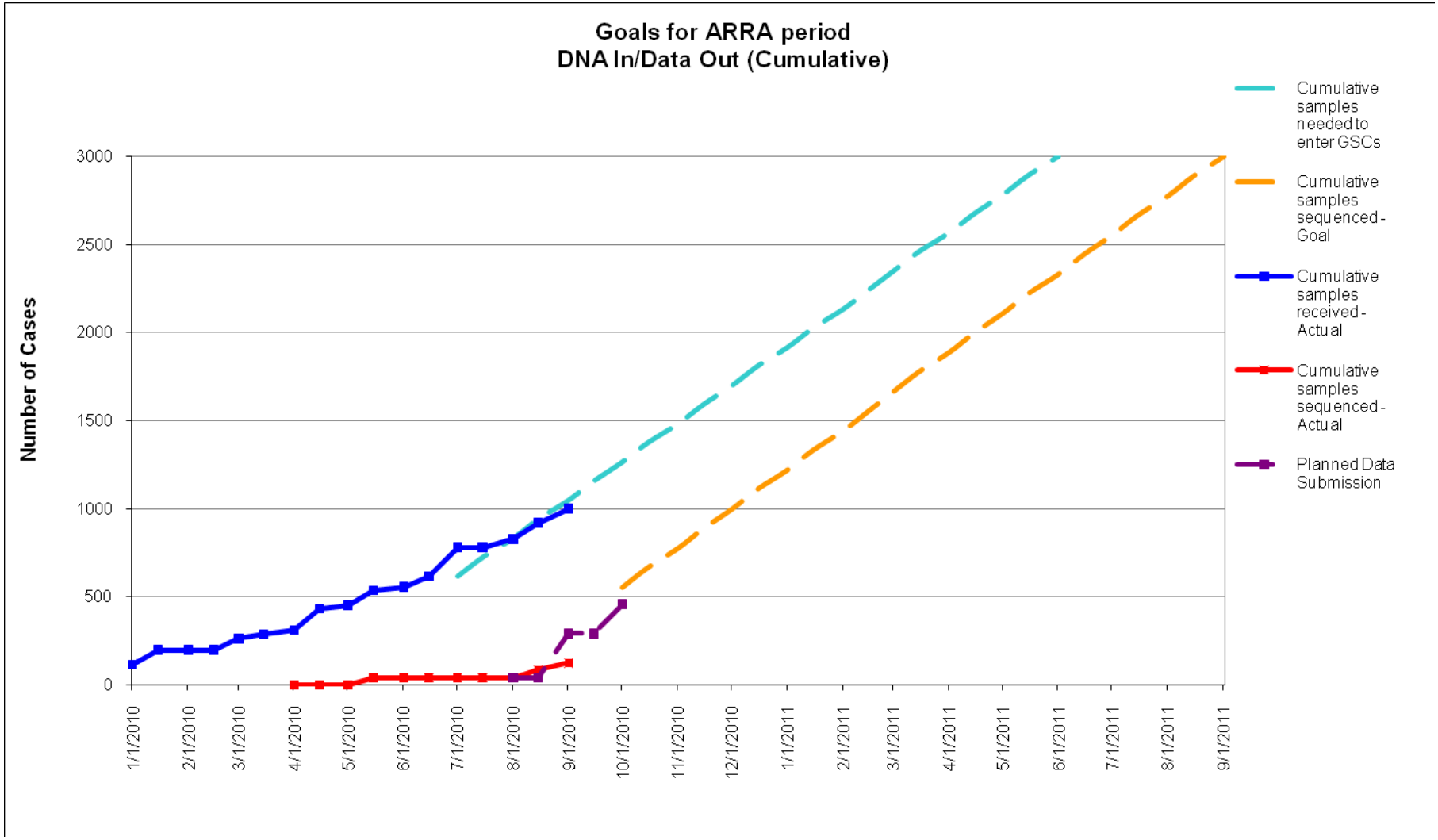
THE CANCER GENOME ATLAS

- USA
  - TCGA projects
- Canada
  - pancreatic*
  - prostate
- Australia
  - pancreatic*
  - ovarian
- China
  - stomach
- EU/France
  - renal carcinomas
- EU/United Kingdom
  - breast cancers*
- France
  - breast cancers*
  - hepatic (alcohol-associated)
- Germany
  - pediatric brain cancers
- India
  - oral
- Italy
  - rare pancreatic types
- Japan
  - hepatic (virus-associated)
- Spain
  - chronic lymphocytic leukemia
- United Kingdom
  - breast cancers*

*Collaborative

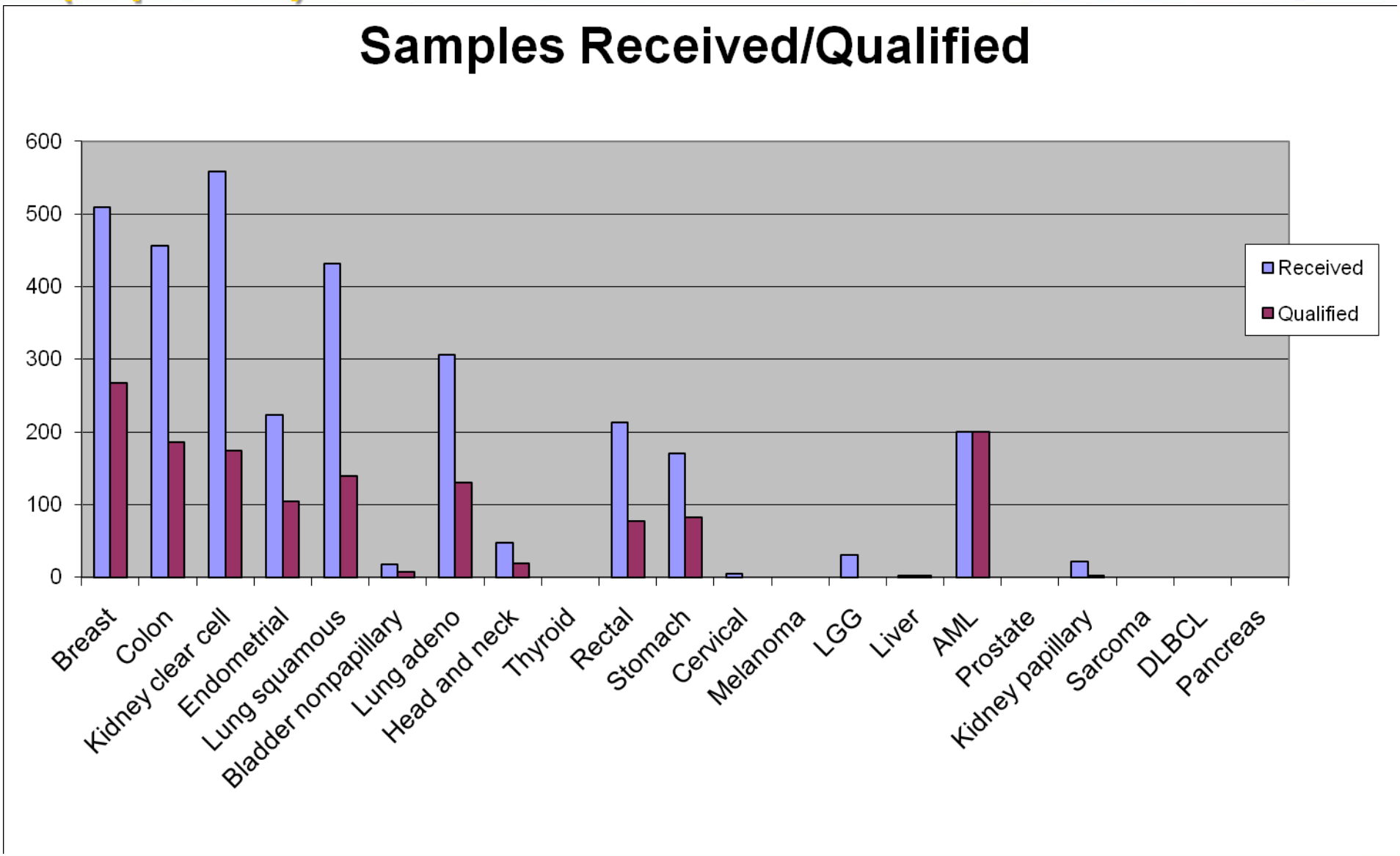
Goals for ARRA Period

THE CANCER GENOME ATLAS



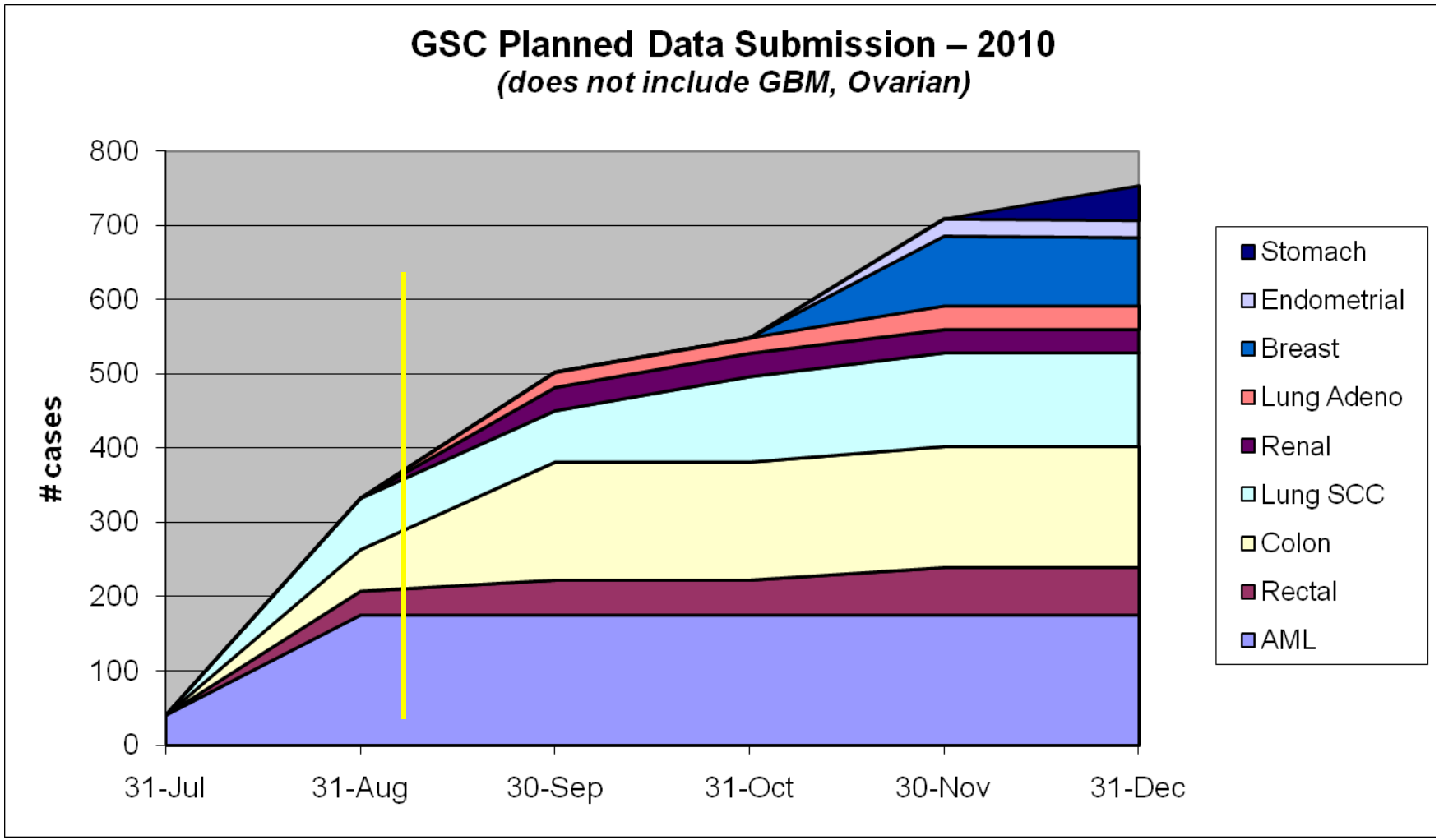
TCGA Sample Accrual
(Sept 2010)

THE CANCER GENOME ATLAS



TCGA Sequence Data Submission

THE CANCER GENOME ATLAS



TCGA to date t

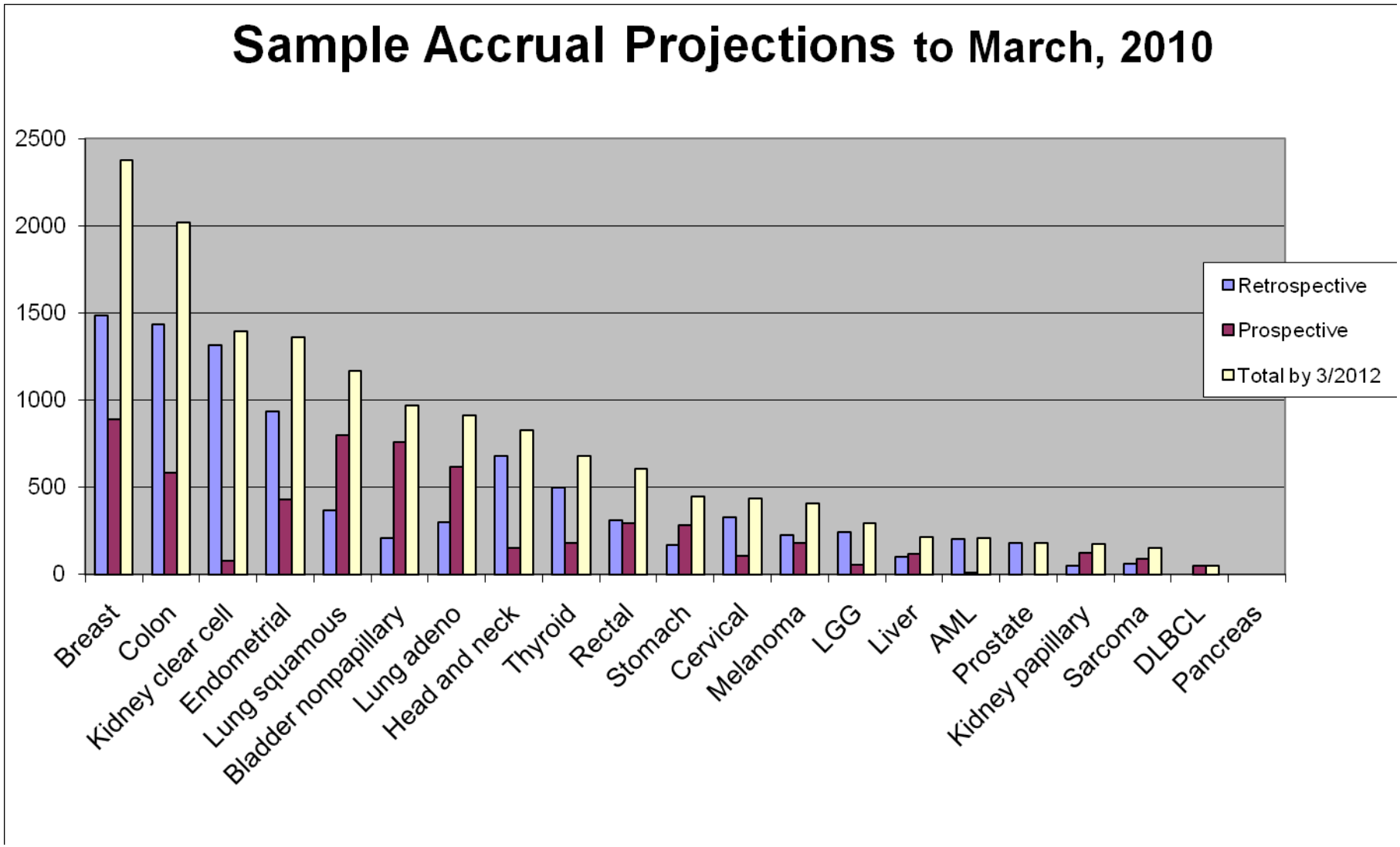
THE CANCER GENOME ATLAS

Tumor Type	GCC assays	Whole Exomes	Whole Genomes
Ovarian	560	434 86 in progress	10 17 in progress
AML	162 39 in progress	15 135 in progress	26 29 in progress
Colon	103 41 in progress	52 51 in progress	0
Rectal	50 17 in progress	0 67 in progress	0
Breast ductal	0 233 in progress	0 186 in progress	0
Lung adeno	21 74 in progress	0 95 in progress	0
Lung scc	69 45 in progress	0 114 in progress	0
Endometrial	0 70 in progress	0 70 in progress	0
Renal	32	0 32 in progress	0
Gastric	0 82 in progress	0 82 in progress	0

http://cancergenome.nih.gov/dataportal

TCGA Sample Accrual
(through March 2011)

THE CANCER GENOME ATLAS



TCGA Sample Criteria (Pilot)

THE CANCER GENOME ATLAS

- Primary tumor only
- Snap frozen
- ~ 200 mg
- No more than 20% necrosis ; ≥ 80% tumor cells
- Normal tissue: Blood (buffy coat/white cells); adjacent normal tissue or buccal cells; or ≥ 13µg high-quality DNA
- All “Tier One” Clinical Data Elements (15 or more)
- Treatment naïve

Challenges for TCGA Production

- **Sample criteria**
  - **Tumors for which pre-treatment is standard of care**
  - **Tumors of lower purity**
  - **Use of adjacent tissue as “germ line” comparison**

Challenges for TCGA Production

- Sample criteria
  - Tumors for which pre-treatment is standard of care
  - Tumors of lower purity
  - Use of adjacent tissue as “germ line” comparison
- **Multiple concurrent projects**
- **Project length**

Organization of TCGA Sequencing

THE CANCER GENOME ATLAS

- ***GBM and ovarian pilot projects*** **– samples were apportioned among the 3 GSCs**

Organization of TCGA Sequencing

THE CANCER GENOME ATLAS

- *GBM and ovarian pilot projects* – samples were apportioned among the 3 GSCs

- *AML and colon* – designated primary and secondary GSCs; 10 exomes and 1 whole genome cases to be done by both centers
  - AML – Wash U / Broad Institute
  - Colon – Baylor / Wash U

Organization of TCGA Sequencing

THE CANCER GENOME ATLAS

- *GBM and ovarian pilot projects* – samples were apportioned among the 3 GSCs

- *AML and colon* – designated primary and secondary GSCs; 10 exomes and 1 whole genome cases to be done by both centers
  - AML – Wash U / Broad Institute
  - Colon – Baylor / Wash U

- ***Subsequent projects* – to be assigned to a single GSC for all data generation, validation, and analysis; 10 exomes to be duplicated in a second GSC for QA**
  - **Broad Institute: lung adeno, lung squamous, gastric**
  - **Wash U: breast projects, endometrial**
  - **Baylor: rectal, kidney**