CENTER FOR CANCER RESEARCH

CCR Bioinformatics Resources

Dr. Eric Stahlberg

Director – CCRIFX Bioinformatics Core

March 25, 2014

Eric.Stahlberg@nih.gov

Frederick National Laboratory for Cancer Research

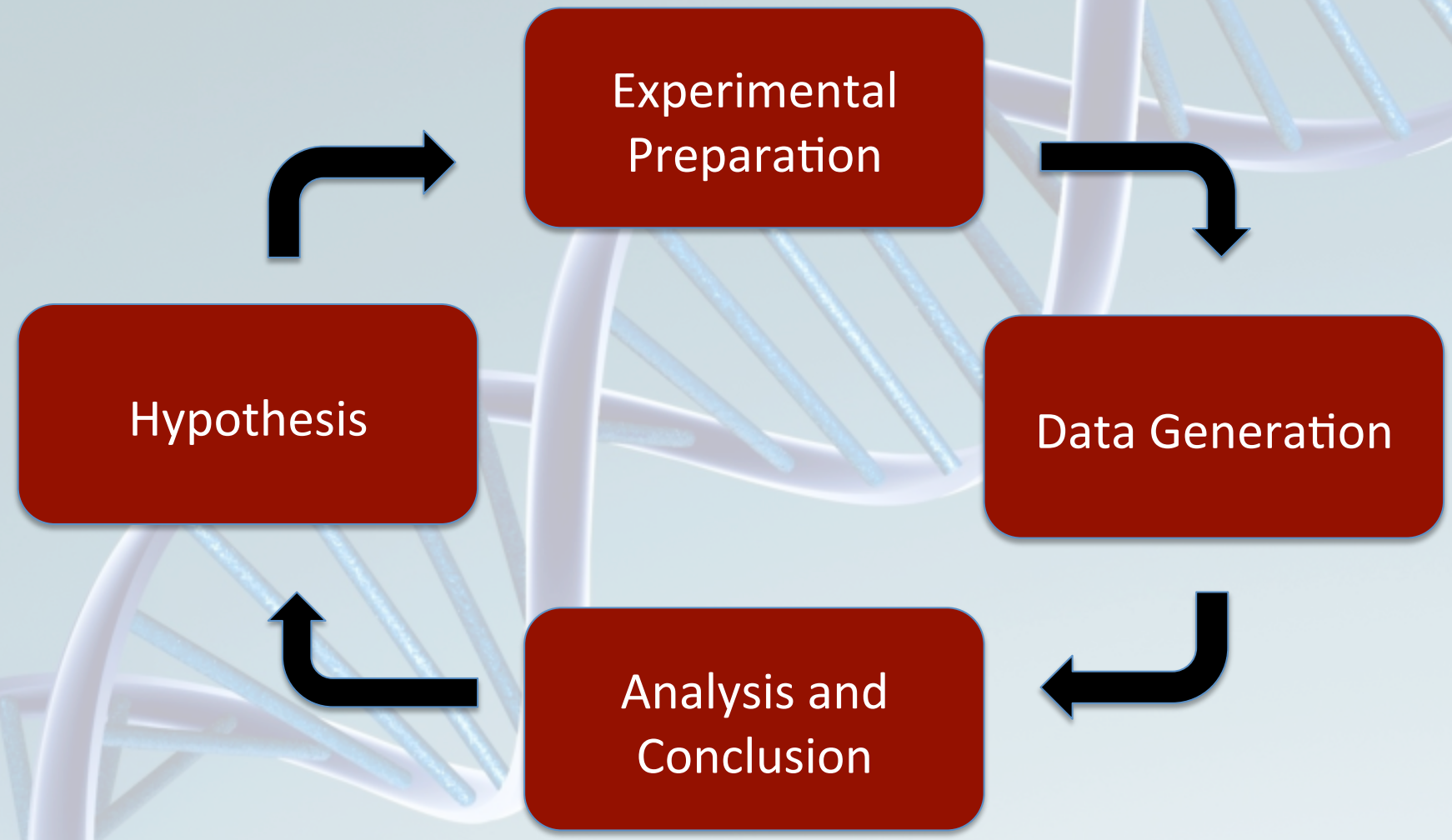
leidos

Leidos Biomedical Research, Inc.

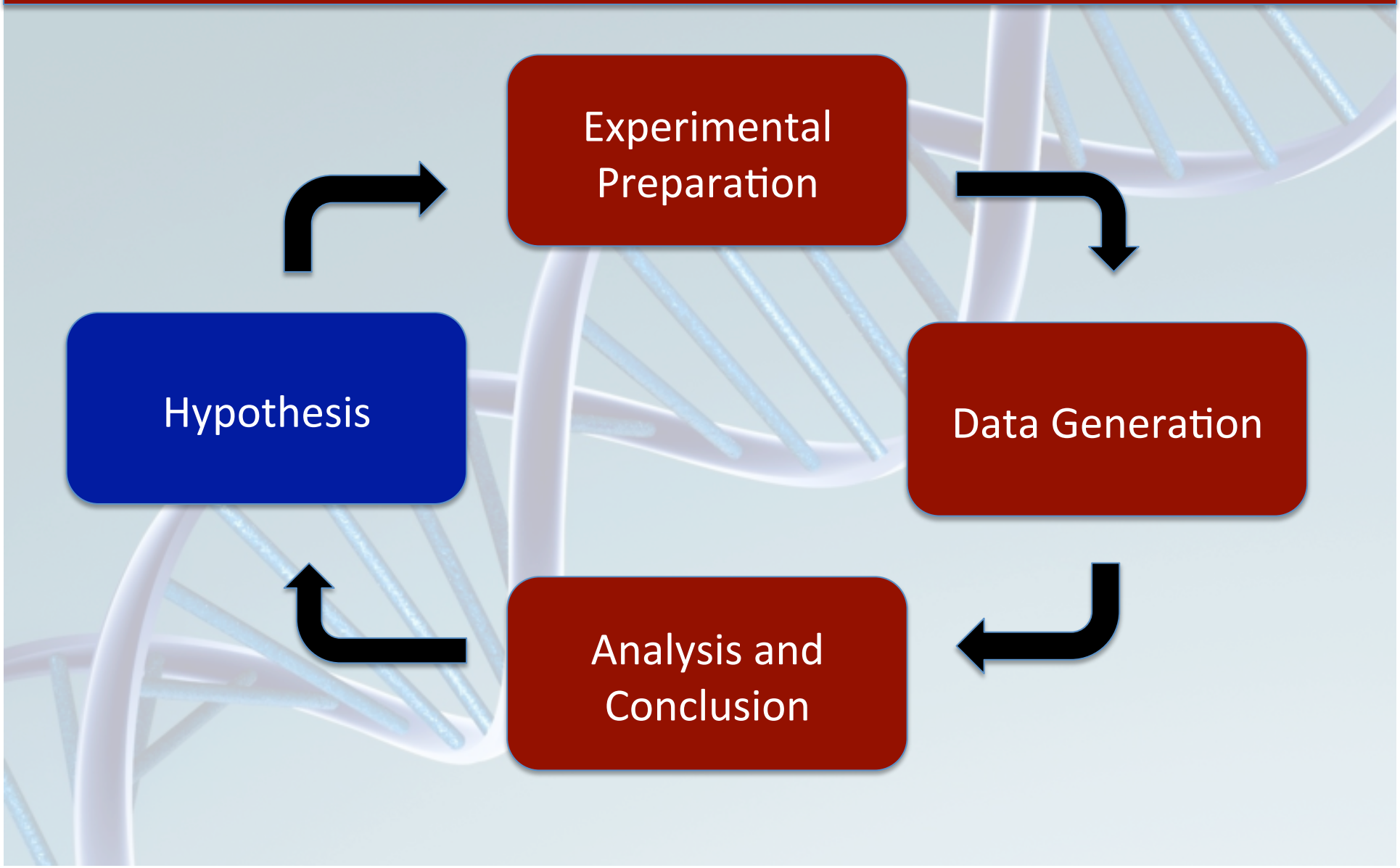
Resource Overview

- **OSTR** – Office of Science and Technology Resources
  - Starting point for most needs
  - Contact: **Dr. David Goldstein**
- **CCRIFX** – CCR Informatics Core
  - Starting point for bioinformatics analysis and experimental design questions
  - Contact: **Dr. Eric Stahlberg or Dr. Natalie Abrams**
- **CCR SF** – CCR Sequencing Facility
  - Where to get sequencing done
  - Contact: **Bao Tran**
- **ABCC** – Advanced Biomedical Computing Center (Leidos/FNL)
  - Source for more general questions (web resource development, image and visualization, mathematical modeling, proteins and structure analysis, etc.)
  - Contact: **Dr. Jack Collins**
- **CIT** – NIH Center for Information Technology
  - Where to get biowulf and helix accounts
- **BTEP** – Bioinformatics Training and Education Program
  - Bioinformatics software training for CCR scientists
  - Contact: **Dr. Maggie Cam or Dr. Peter FitzGerald**
- **CREX** – CCR Research Exchange
  - Where to go to find new and available resources and new contacts

Scientific Process Schematic



Scientific Process Schematic

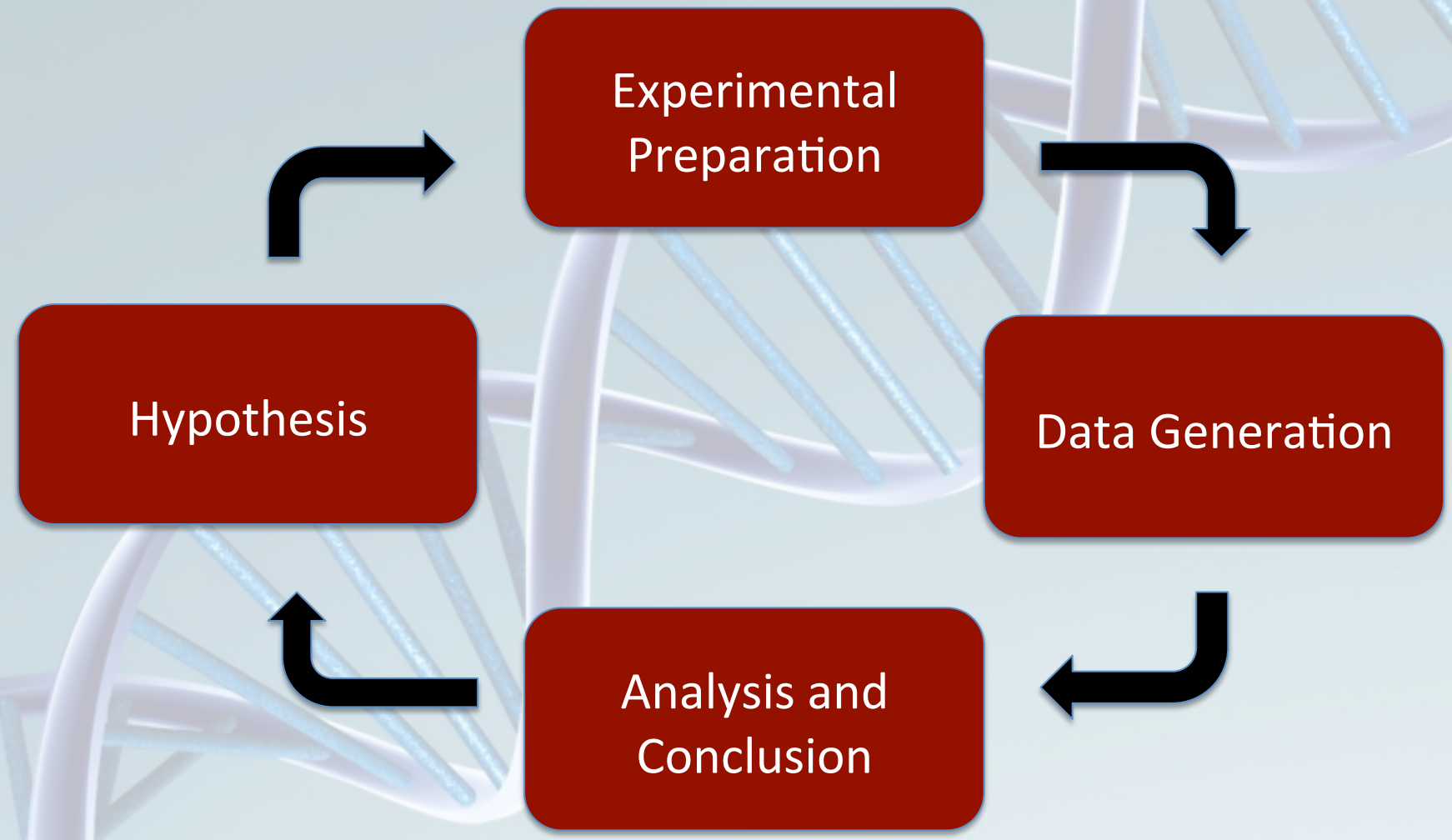


Hypothesis

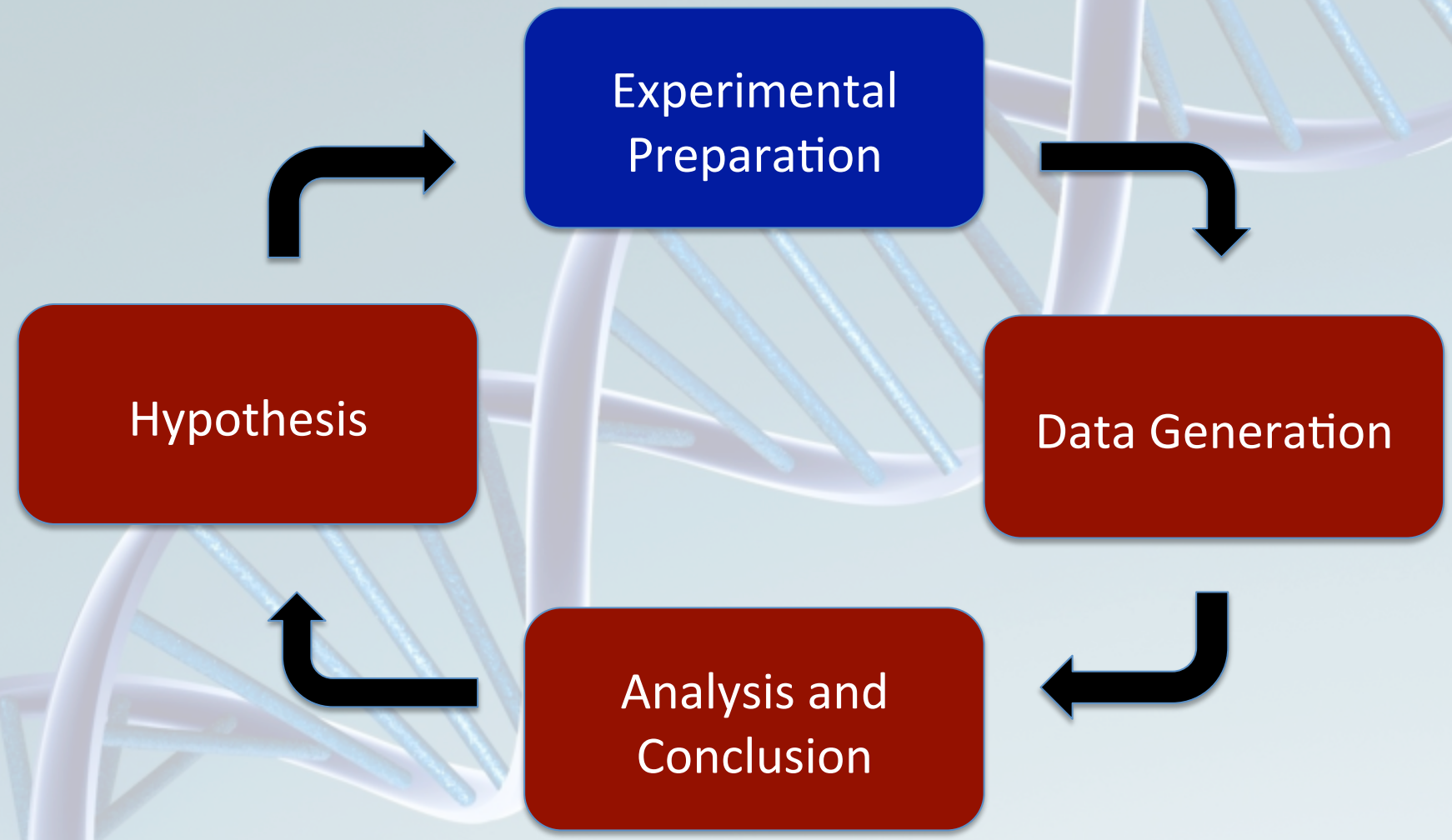
- What would you like to know?
- What would you like to learn?
- Defining scientific context
- Driving scientific questions

As these become clear, where to go next?

Scientific Process Schematic



Scientific Process Schematic



Experimental Preparation

- What technologies are available?
- Which approach should I use?
- Where is next generation sequencing done?
- How do I get started?
- Who does microarray analysis?
- What elements help assure a good analysis?
- Who has advice and insight into options?

Resources

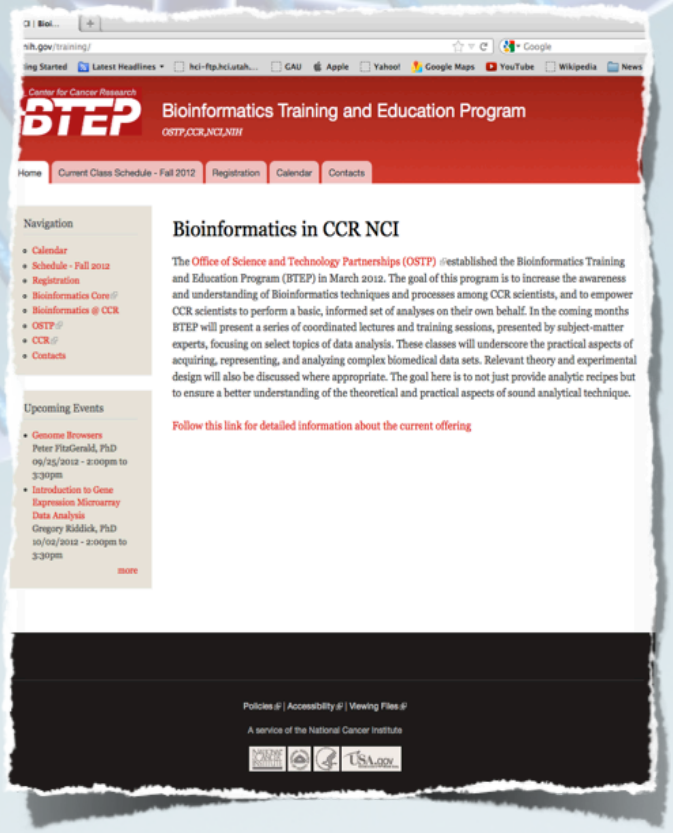
- Training opportunities in BTEP
- Experimental design consultations
  - Simple access – http://ccrifx.cancer.gov
  - Access CCRIFX core AND CCR bioinformaticians
- Technology and application options

Center for Cancer Research

BTEP

Bioinformatics Training and Education Program

The Office of Science and Technology Resources (OSTR) established the Bioinformatics Training and Education Program (BTEP) in March 2012.



http://bioinformatics.nci.nih.gov/training/ *(Only accessible via the NIH network)*

Center for Cancer Research

BITEP

Bioinformatics Training and Education Program

The Goal: To increase the awareness and understanding of Bioinformatics techniques and processes among CCR scientists, and to empower them to perform a basic set of analyses on their own behalf.

The Process: Present a series of coordinated lectures and training sessions on select topics, as part of a predefined program.

Target Audience: Initially at least, the classes are aimed at Post-Doc level biomedical researchers with no special training in bioinformatics.

Center for Cancer Research

BTEP

Bioinformatics Training and Education Program

The BTEP program offers training in three different formats

- **Lectures** – these are typically 1-2 hours long and present a detailed exploration of a specific bioinformatics topic. *(16 Classes)*

- **Hands-on classes** - consisting of either a single or a pair of 3 hour sessions these offerings cover the detailed working of a particular software package. *(13 Classes)*

- **Workshops** – these 2-day courses include both lecture and hands-on components, and cover the basic concepts and practical aspects of a specific type of data analysis (e.g. ChIPSeq). Each class covers experimental design, statistical analysis, pitfalls and general workflows and includes the use of both commercial and open source software. *(3 Classes)*

Experimental Design

- ❑ Clarify experimental objective
- ❑ Criteria for successful analysis
- ❑ Navigate technology and application choices
- ❑ Feasibility of approach
- ❑ Standards, guidelines and recommendations
- ❑ Controls and numbers of replicates
- ❑ Cost effectiveness

- ❑ To get started
  - Visit CCRIFX website (http://ccrifx.cancer.gov)
  - Submit a brief request
  - No cost support

Technology and Application Selection

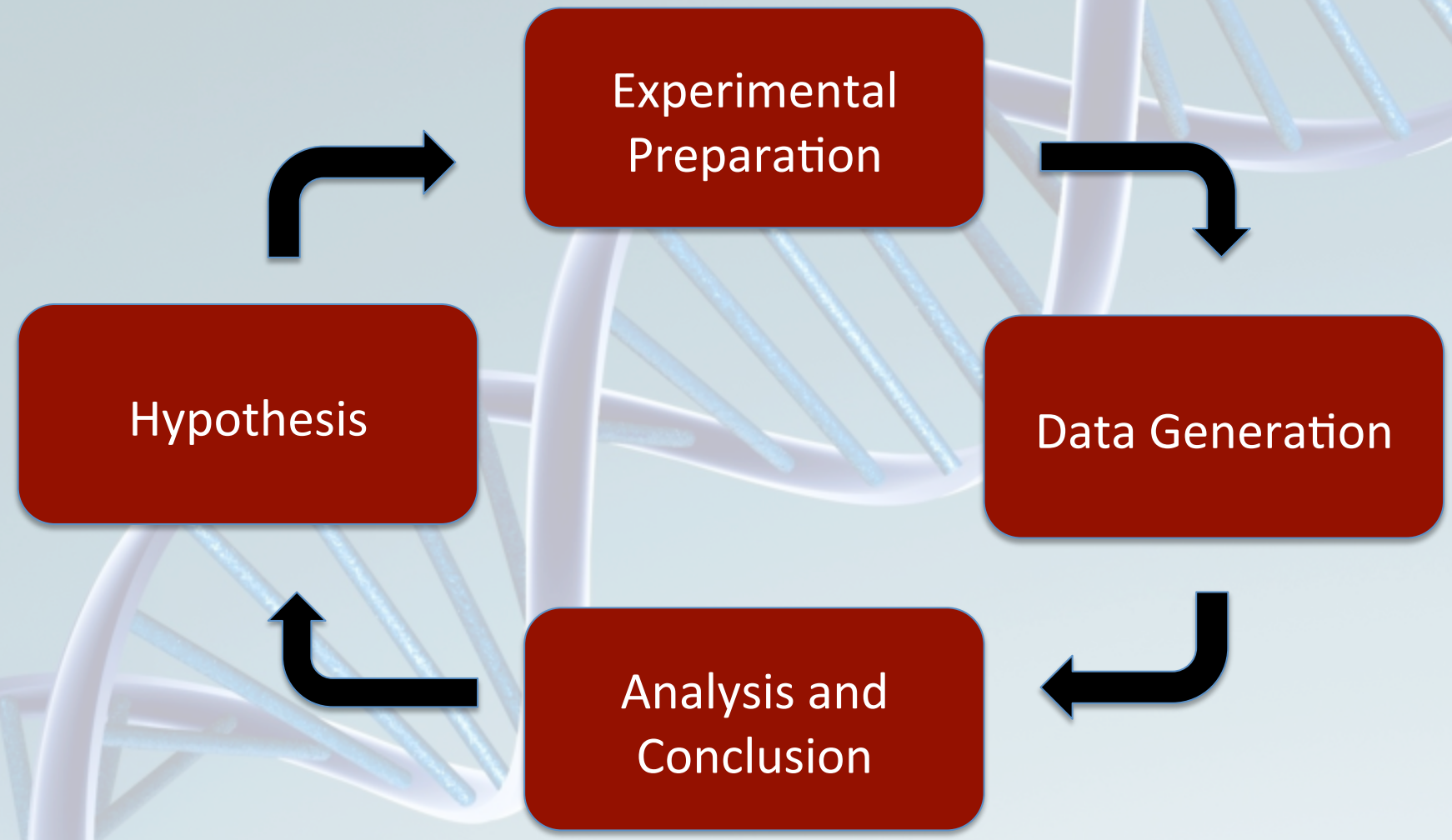
Technology Choices

- Illumina Sequencing IGA IIx,
- Illumina HiSeq 2000,
- Illumina MiSeq
- Pacific Biosciences Pacbio RS II
- Microarray
- SNP array
- ChiP-Seq
- RNA-Seq
- Exome sequencing
- Whole genome sequencing
- Etc.

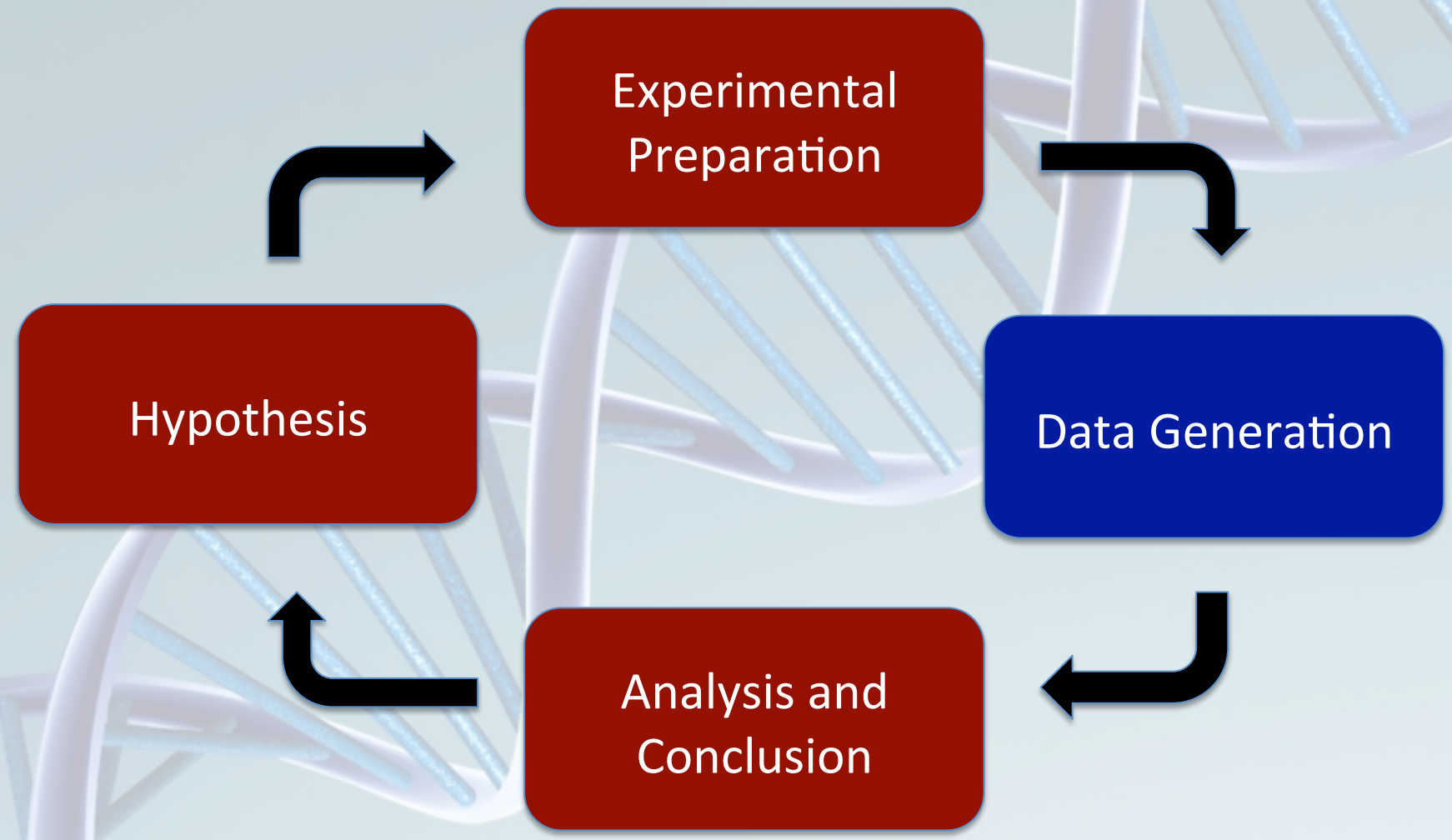
….. and Application Choices

- Differential Expression
- DNA/RNA-Protein Interactions
- Variant Calling and Impact Analysis
- Gene Fusions
- Epigenetic Investigations
- Statistical approaches
- Transcript Profiling and Discovery
- Genome-wide Methylation Study
- Targeted Exome or Custom Enrichment Sequencing
- Always expanding

Scientific Process



Scientific Process



Data Generation

- Where do I send samples?
- How much does it cost?
- What machines are available?
- How long will it take?
- How do I get the data back?
- How do I know if the data is high quality?
- What data will be available when complete?

CCR Sequencing Facilities

The primary mission of CCR-Sequencing Facility (CCR-SF) is to utilize high-throughput sequencing technologies to enrich cancer research and ensure that the NCI community can remain at the leading edge of next-generation sequencing technology.

- Located at the Advanced Technology Research Facility (ATRF), Frederick, MD
- Currently Operate:
  - Illumina Sequencing: (2) IGA IIx, (3) HiSeq 2000, (1) MiSeq
  - Pacific Biosciences: (1) Pacbio RS II
- During this past fiscal
  - provided Illumina sequencing and support to 49 NCI-Investigators
  - processed and sequenced more than 1,658 samples
- *Major Applications on the Illumina Sequencing platform include:*
  - **DNA/RNA-Protein Interactions**
  - **Transcript Profiling and Discovery**
  - **Genome-wide Methylation Study**
  - **Targeted Exome or Custom Enrichment Sequencing**

http://ncifrederick.cancer.gov/atp/genetics-and-genomics/

Sequencing Request – Self Analysis

- Self Analysis Request
  - Step 1. Prepare experimental details
  - Step 2. Submit sequencing request – indicating self-analysis
    - Submit https://ostr.cancer.gov/sequence/request
    - **Indicate self analysis on sequencing request form**
  - Step 3. Review/approval/consultation
  - Step 4. Submit samples
  - Step 5. Data made available to you from sequencing facility

Sequencing Request - Collaborative

- Collaborative Analysis Request
  - Step 1. Prepare experimental details
  - Step 2. Secure analysis request number (if not already done)
    - Submit request with quick overview at http://ccrifx.cancer.gov
  - Step 3. Submit sequencing request
    - Submit https://ostr.cancer.gov/sequence/request
    - **Request number from Step 2 used to link sequencing to analysis**
  - Step 4. Review/approval/consultation
  - Step 5. Submit samples
  - Step 6. Collaborative analysis starts when data is ready

Online Sequencing Request

Sequence Request Form | OSTR

CCR IFX :: Project Details | Sequencing Facility (SF) | Advan... | SF Services – Project design, su... | Sequence Request Form | OSTR

https://ostr.cancer.gov/sequence/request

Center for Cancer Research

OSTR

Office of Science and Technology Resources

Resources Partnerships Bioinformatics Developing Technologies About Subsidy Program

Sequence Request Form

Home » Sequence Request Form

If you are planning on using the CCRIFX bioinformatics core for the analysis of this project, please submit an IFX project request prior to submitting this form.

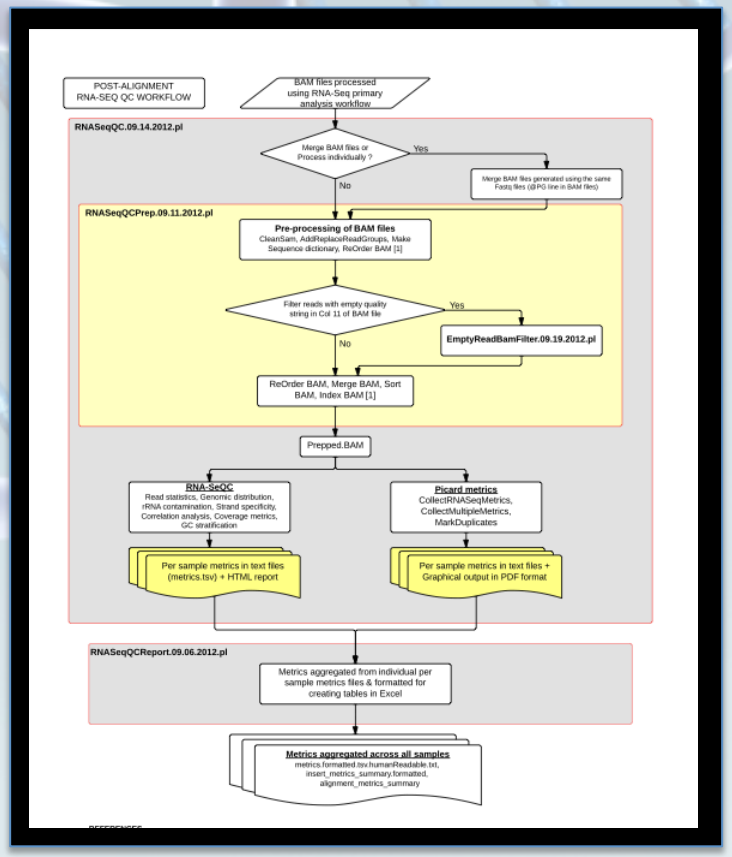
Bioinformatics Support *

- I have an active project with the CCRIFX bioinformatics core for the analysis involving this data
- I am expecting to complete my own analysis of this data or I am collaborating with a Bioinformatician

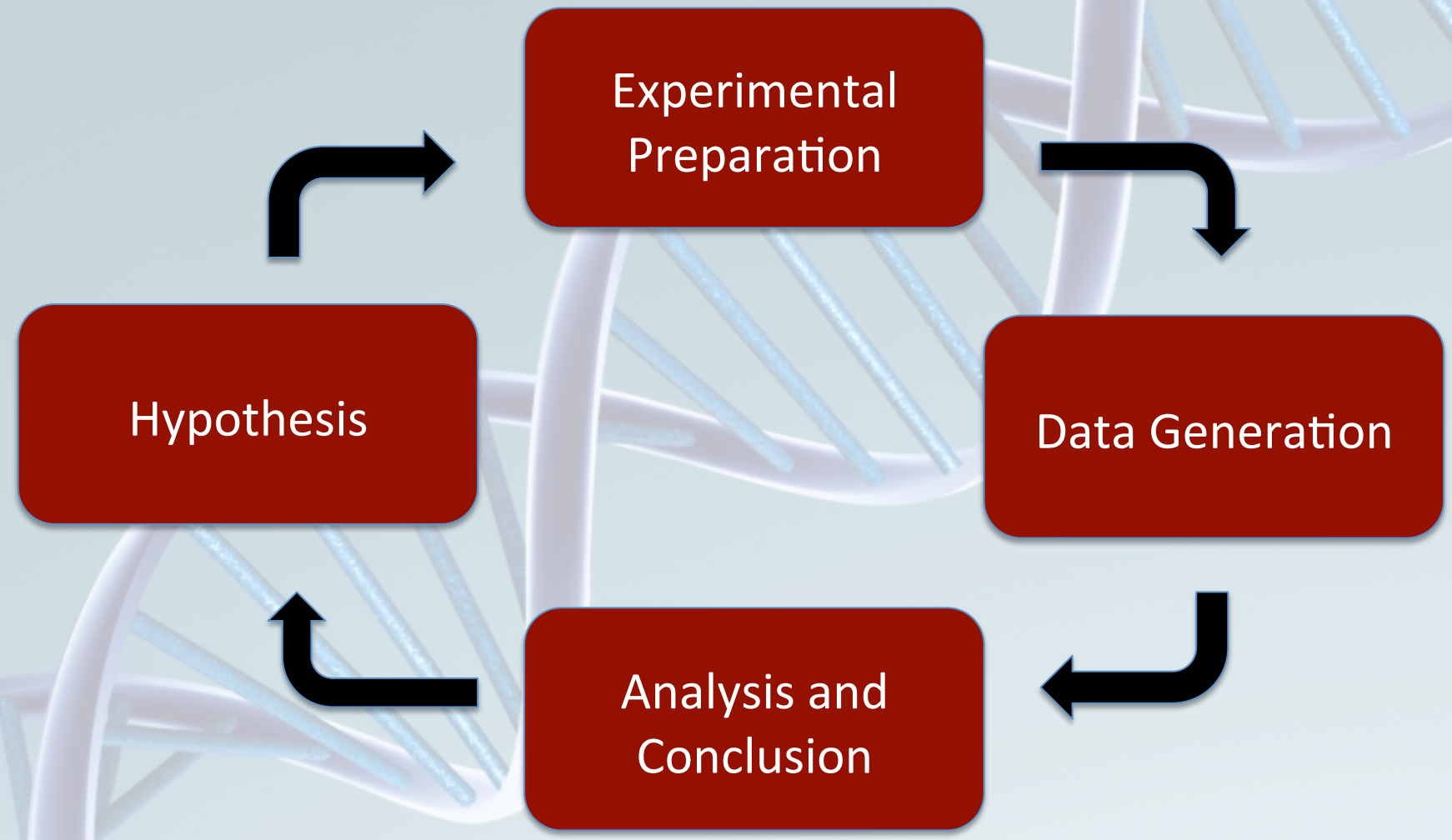
Next >>

Quality and Primary Analysis

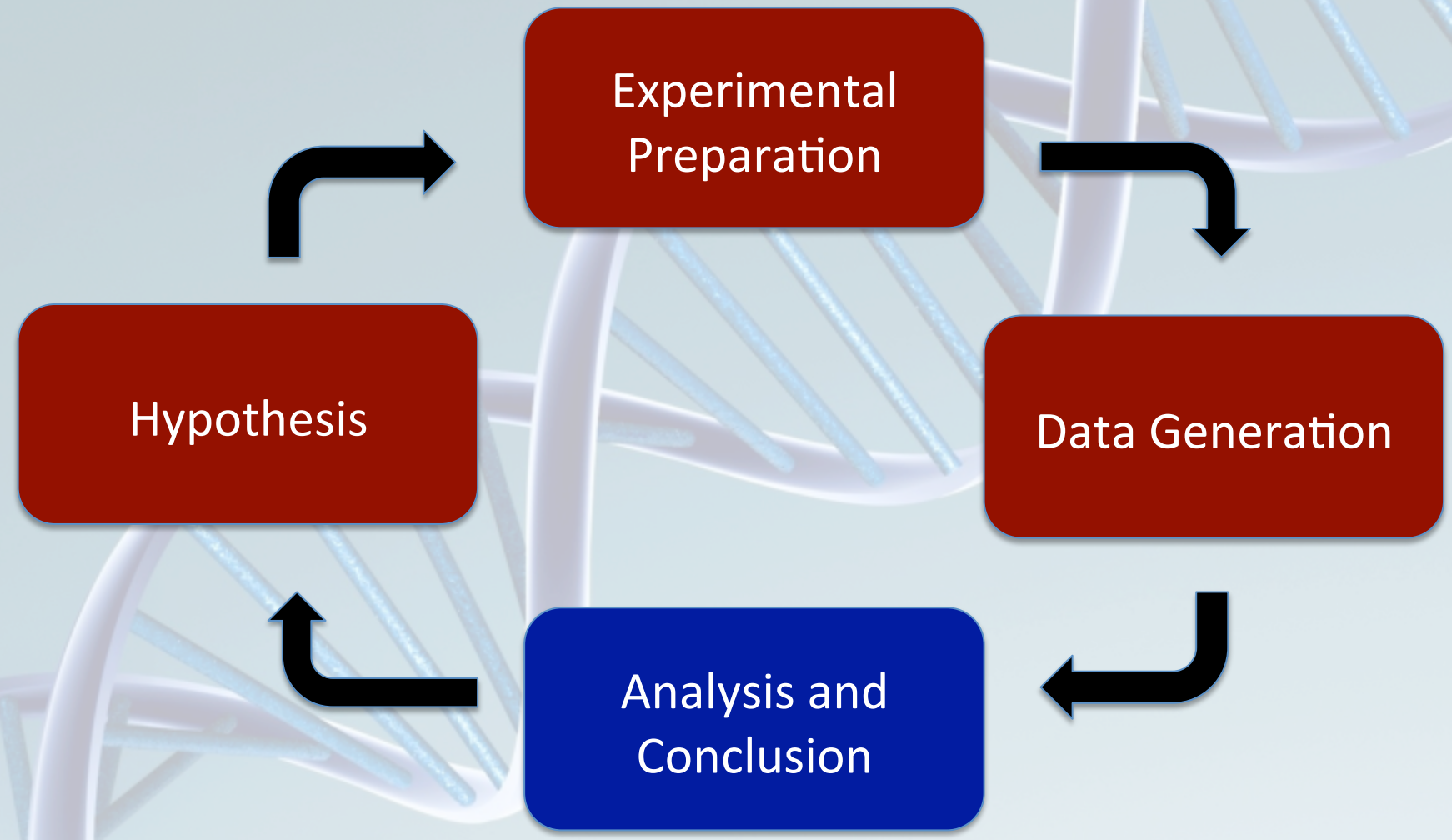
- CCR SF IFX Delivers
  - Data quality management
    - Feedback loop to investigator
    - Online quality reports
    - Robust QC workflows
  - Original data files
    - BAM format
    - FASTQ format
    - Helix accounts (CIT)
    - Collaborative ftp download (hosted by CIT)
  - Selected secondary analysis



Scientific Process



Scientific Process



Analysis and Conclusion

- I have all of this data, what's next?

CCR Bioinformatics

- CCR Bioinformatics Resources
  - CCR Research Exchange (CREX)
  - CCRIFX Bioinformatics Core
  - Bioinformatics Training and Education Program (BTEP)
  - Compute nodes on NIH Biowulf
  - Analysis software and systems
  - Much more!

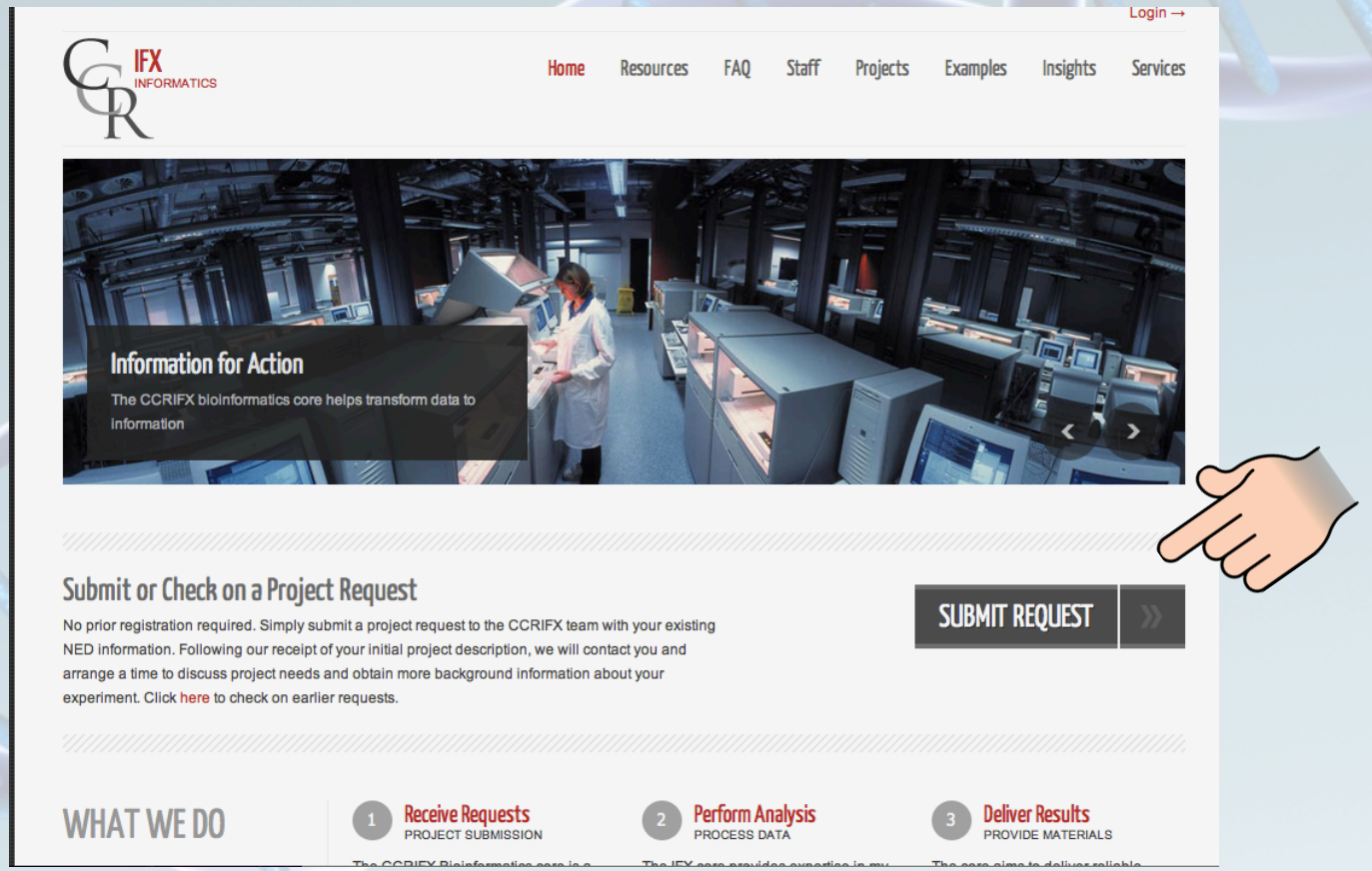
Visit https://ostr.cancer.gov

CCRIFX Bioinformatics Core

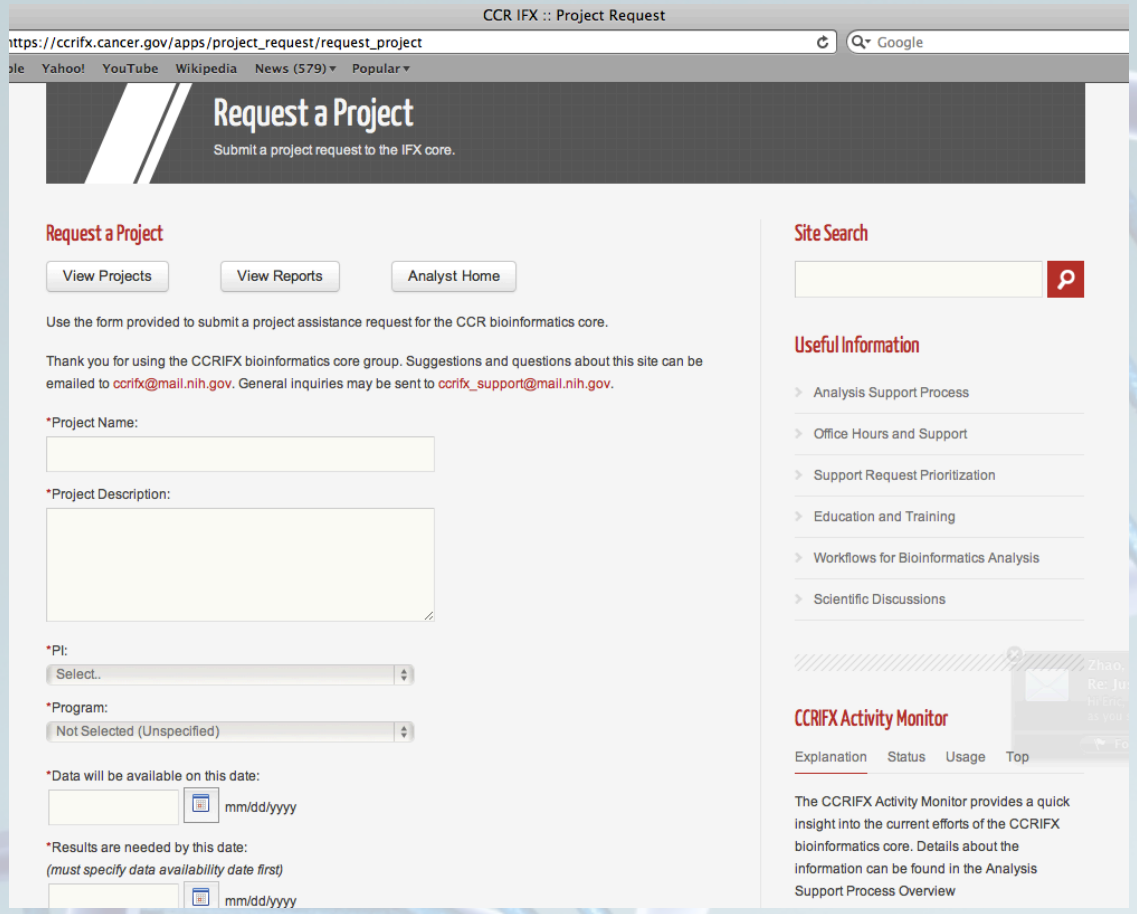
- Started 2011 as Collaborative Bioinformatics Analysis Core
  - Collaborative effort between CCR and Leidos Biomedical Research Inc.
  - Part of the Advanced Biomedical Computing Center
  - Working together to provide needed bioinformatics experience across CCR
- Providing **no-cost** and collaborative
  - Bioinformatics analysis (secondary, custom, integrated)
  - Experimental design
  - Focused/customized training
- Experience with a wide variety of bioinformatics software used with over 100 unique projects
- Support delivered to over 200 scientists and 97 labs in just over 3 years
- Predominant Technologies
  - Chip-Seq, RNA-Seq, Exome-Seq, miRNA-Seq, microarray, etc.
- See http://ccrifx.cancer.gov website for more details

CCRIFX Bioinformatics Core

http://ccrifx.cancer.gov



Continue or Submit New Request



Items Needed

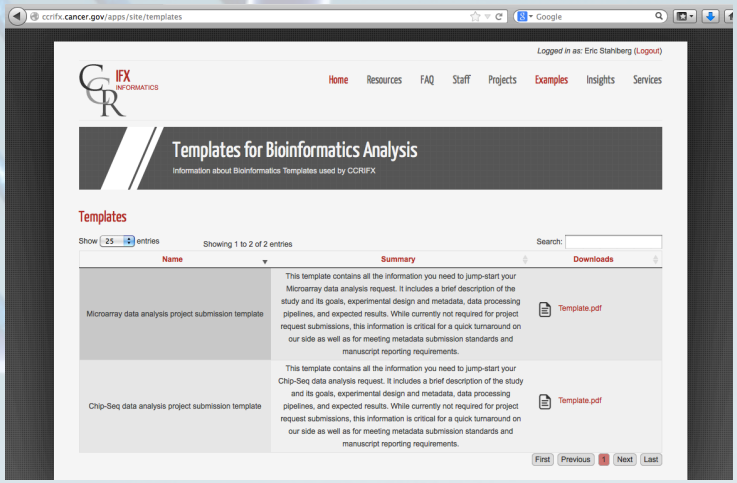
- Project name
- Project description text
- PI information
- Date data is available
- Date results desired by

If continuing analysis from experimental design request, nothing more likely needed.

http://ccrifx.cancer.gov

CCRIFX Resources

- Additional Resources at CCRIFX site
  - Templates for different analysis types
    - Microarray, ChiP-seq
  - Experimental Design Guidance
  - Presentations and Tutorials
  - Workflows and documentation
    - Quality Control
    - RNA-Seq analysis
  - Frequently Asked Questions
  - Pointers to other resources



Center for Cancer Research

BTEP

Bioinformatics Training and Education Program

A non-exhaustive list of the topics covered includes the following:

- **Gene Expression -** Microarray and RNASeq
- **Pathway Analysis -** IPA, GeneGo, GSEA, Pathway Studio
- **Visualization** - Genome Browsers, Integrative Genomics Viewer (**IGV**)
- **Transcription Factor Binding -** ChIPSeq, Genomatix,
- **Variant Discovery -** Nexus Copy Number, SNP calling
- **Cancer Genomics** - TCGA, CBio
- **Commercial Software Packages**

  Partek, Genomatix, CLCBio, Geneious,IPA,Nexus, Pathway Studio

Center for Cancer Research

BITEP

Bioinformatics Training and Education Program

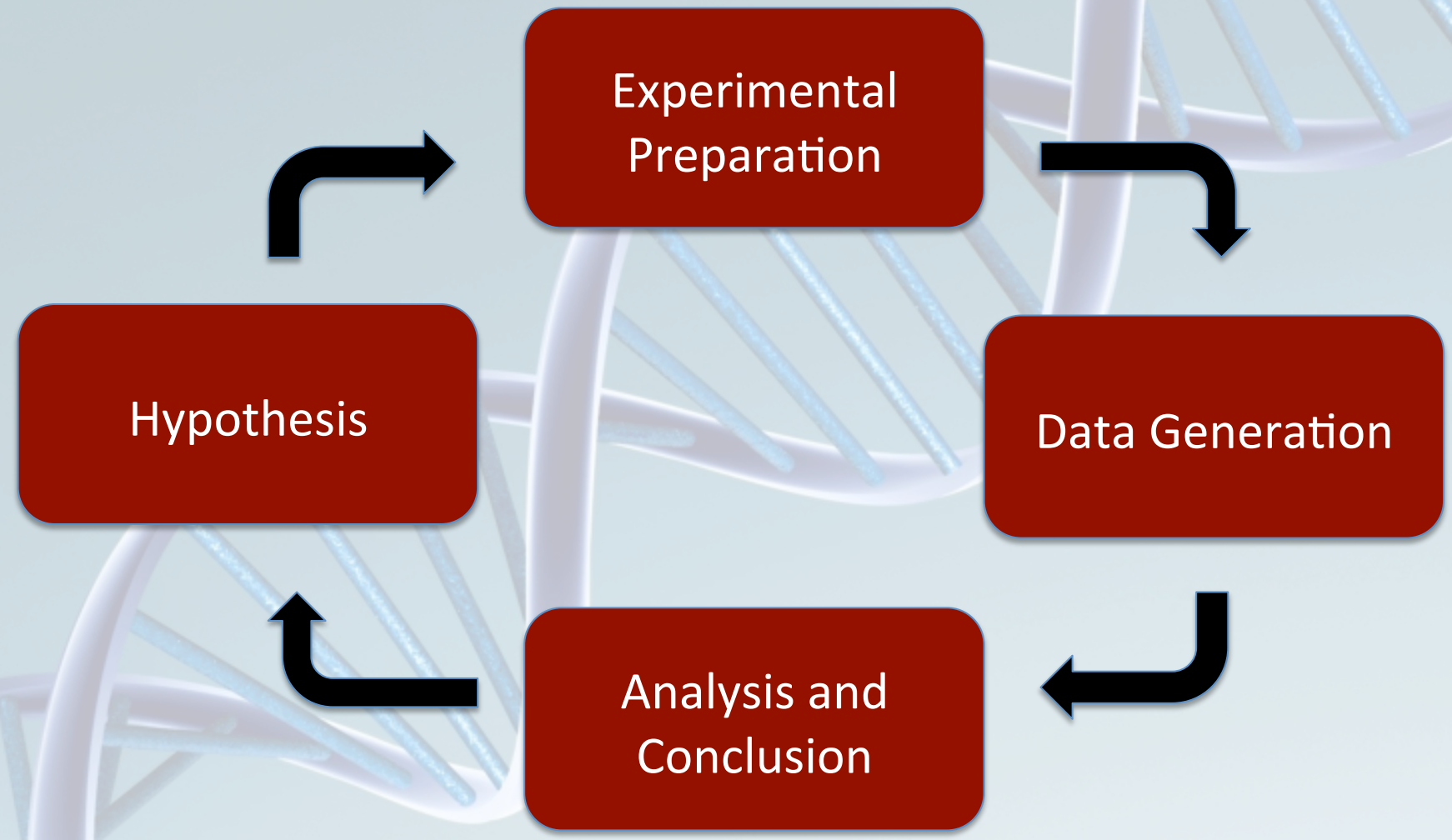
Future Directions:

- **FAQs** for most common problems and software quick starts
- **Forums** online communities for those seeking assistance
- **WEB-based applications** for common tasks
- The introduction of **advanced topics**
- Establish archives of **video tutorials**

Comments and suggestions to: NCIBTEP@mail.nih.gov

http://bioinformatics.nci.nih.gov/training/ *(Only accessible via the NIH network)*

Scientific Process



Abundant Resources for Success

YOUR RESEARCH SUCCESS

Hypothesis

Experimental Preparation

Data Generation

Analysis and Conclusion

CCR

CCR Bioinformatics

ANALYSIS

SOFTWARE

CIT

Experimental Design

CCRIFX

BTEP

ISP

CREX

CCRIFX

Leidos

ANALYSIS

FNL

CCR SF

ABCC

OSTR

Some Pointers

- Visit Office of Science and Technology Resources website
  - https://ostr.cancer.gov
  - Contact: Dr. David Goldstein
- Reach out for assistance in experimental design early
  - http://ccrifx.cancer.gov
  - Contact: Dr. Eric Stahlberg or Dr. Natalie Abrams
- Take advantage of training oportunities
  - http://bioinformatics.nci.nih.gov/training
  - Contact: Dr. Maggie Cam or Dr. Peter FitzGerald
- Secure helix/biowulf accounts for big computing resources
  - http://cit.nih.gov
- Use the request form to help coordinate your NGS support
  - https://ostr.cancer.gov/sequence/request
  - Contact: Bao Tran
- Stay familiar with technology options, they do change
  - http://ncifrederick.cancer.gov/atp/genetics-and-genomics/

Thank You

Bookmark These Sites!

- ❑Office of Science and Technology Resources website
  - •https://ostr.cancer.gov
- ❑CCRIFX Bioinformatics Core
  - •http://ccrifx.cancer.gov
- ❑BTEP Training
  - •http://bioinformatics.nci.nih.gov/training
- ❑NIH Center for Information Technology
  - •http://cit.nih.gov
- ❑Sequencing Request Form
  - •https://ostr.cancer.gov/sequence/request
- ❑Technology Options
  - •http://ncifrederick.cancer.gov/atp/genetics-and-genomics/