Scientific Writing

魏桐
研究院数字化地球所
10/17/2022

华大生命科学研究院
BGI · research

Outline

- About this course, *Research Ethics and Scientific Writing*
  - Course goals, learning objectives, and grading
- Structure in a research article
- Research proposal

Part I
About this course

华大生命科学研究院
BGI · research

BGI华大 Scientific Reading and Presentation (SRaP)

文献阅读与科技写作（Scientific Reading and Presentation）

- 专业选修课，学分2，周学时4 x 8周=32
- 课程介绍：《文献阅读与科研展示》是一门以文献阅读和展示分享相结合的课程。文献阅读是当代科学研究中的重要一环，研究人员通过阅读文献，了解相关科研领域的最新进展，从而有针对性的提出科学问题、设计科研项目、发表科技论文。文献阅读能力是科研人员的基本功，在研究生教育也有重要地位。通过学习本课程，使学生掌握搜集、阅读和综合文献的能力，学会如何分析一篇科技论文，培养好的阅读习惯，逐渐形成对科研学科的判断力。
- 授课方式：45分钟讲授（本人或者嘉宾），150分钟学生分享讨论（启发式提问）。
- 目标：通过学习本课程，学生应该掌握文献阅读的基本方法；了解基因组学研究的背景；熟悉生物信息学的研究方法；养成良好的文献阅读习惯。
- 考核方式：平时成绩占60%，包括考勤（10%）、ppt展示（40%，团队打分，2次机会）、随堂提问（10%）；期末考试，包括项目汇报（30%）、课后问答题（10%）。

SRaP contents

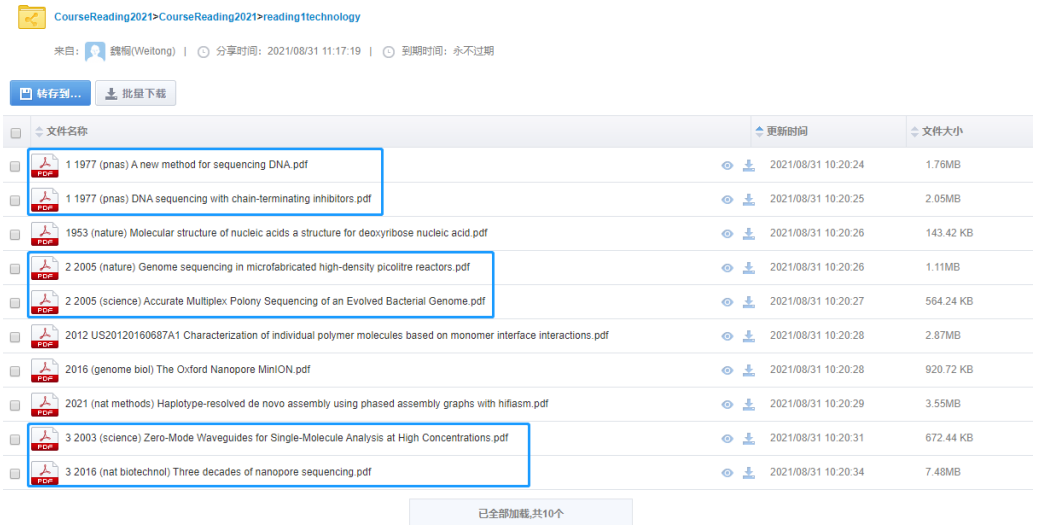
BGI华大

Contents

- Introduction of scientific reading and presentation
- The history of sequencing technology
- Literature about human genomics
- Literature about evolutionary genomics
- Literature about population genomics
- Literature about functional genomics
- Literature about recent omics approaches

Focus on human research

- Final presentation



BGI华大

Learning materials: websites

- Nature milestones on genomics sequencing: https://www.nature.com/immersive/d42859-020-00099-0/index.html
- Human genome project, https://www.genome.gov/human-genome-project
- 1000 Genomes, https://www.internationalgenome.org/
- ENCODE, https://www.encodeproject.org/
- Epigenome Roadmap, http://www.roadmapepigenomics.org/
- The Cancer Genome Atlas, https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga

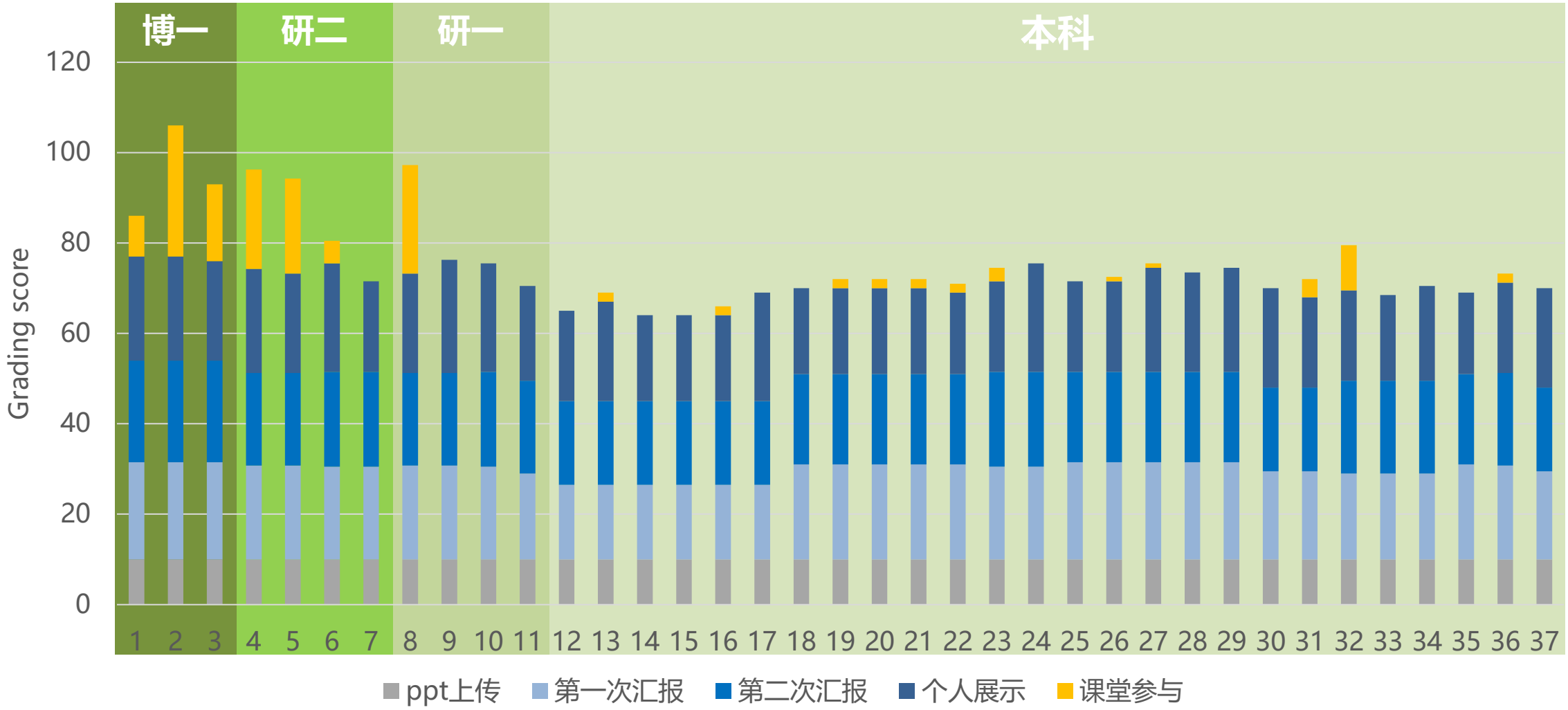
BGI华大

Grading

- Weekly seminar (25% each)
  - ~40-minute talk + 10-minute question
  - Tutor + peer evaluation (on-line survey)
  - Style/format (5%) + presenting (5%)+ background (5%) + results (10%)
- Discussion (25%)
  - Q & A, reporting in 2 days (1-2% each)
- Final presentation (25%)
  - 8-minute talk + 1-2 questions

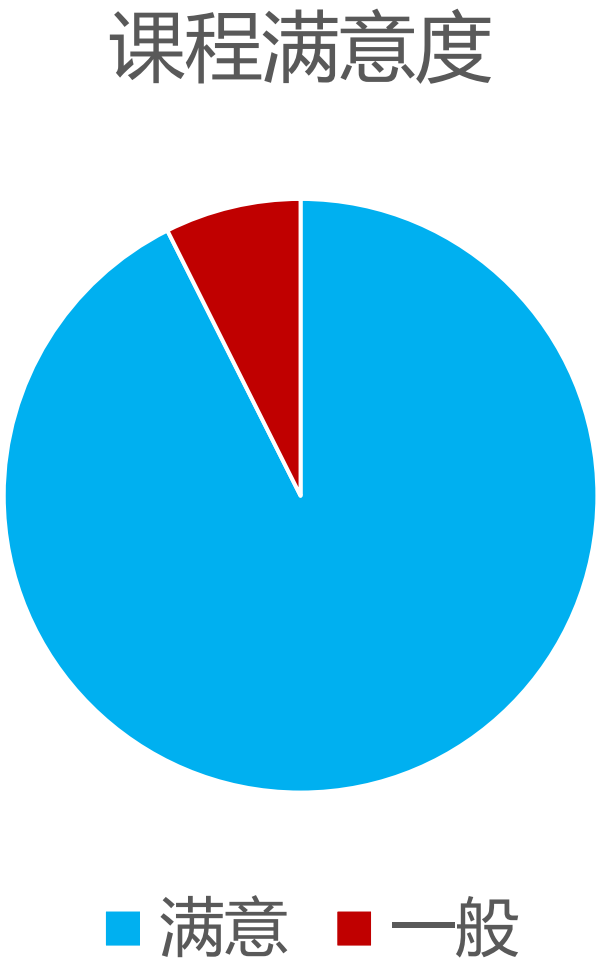
BGI华大

Student score



BGI华大

Student feedback



- 敢于分享
- 从嘉宾分享和老师点评中学到很多
- 提升了**科研思维**与科研能力
- 学会了汇报ppt的制作和讲解，听了很多大佬的讲解
- 在阅读文献，PPT制作和文献展示等方面有较大提升
- 优化文献阅读与展示的**技巧**，系统地接触到各领域**前沿知识**

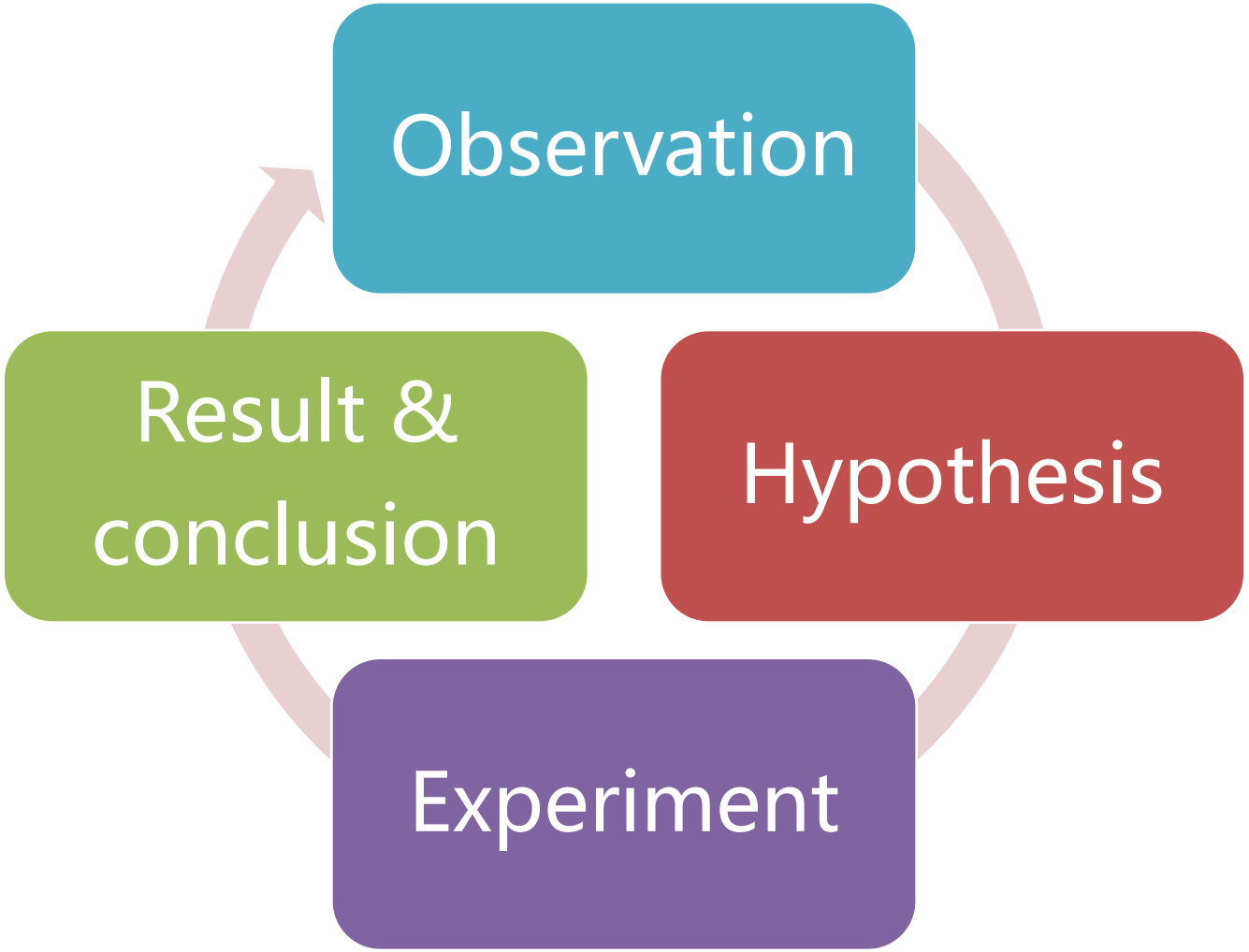
Lessons learned from Reading2021

Learn from **tutor & peer!**

Learn by **practice & discussion!**

BGI华大

Scientific methodology

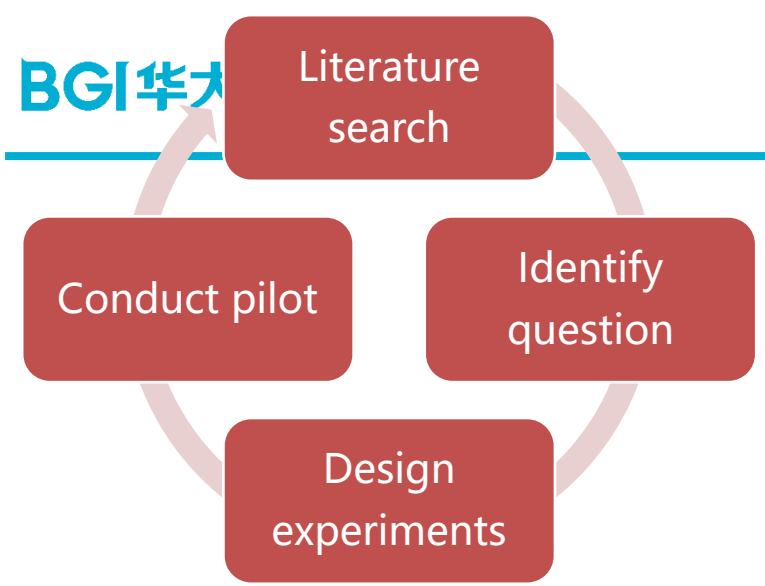


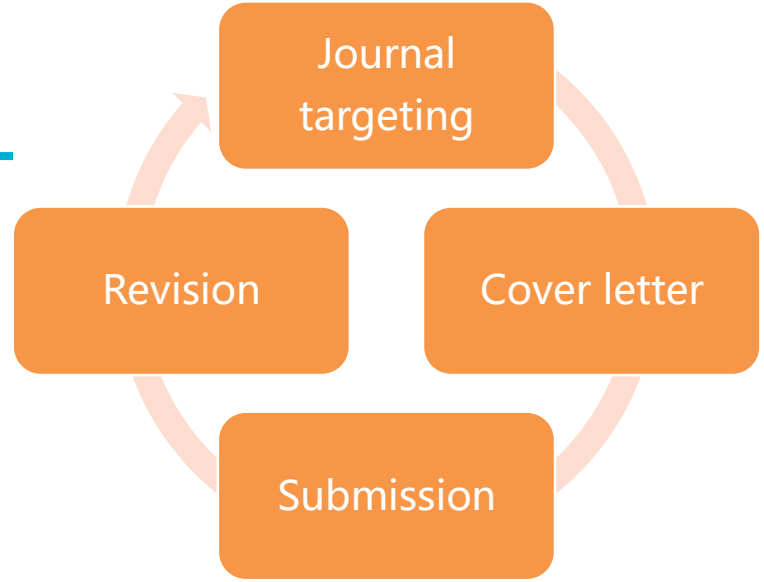
An empirical method of **acquiring knowledge** from observations.

It **applies to all** scientific enterprises.

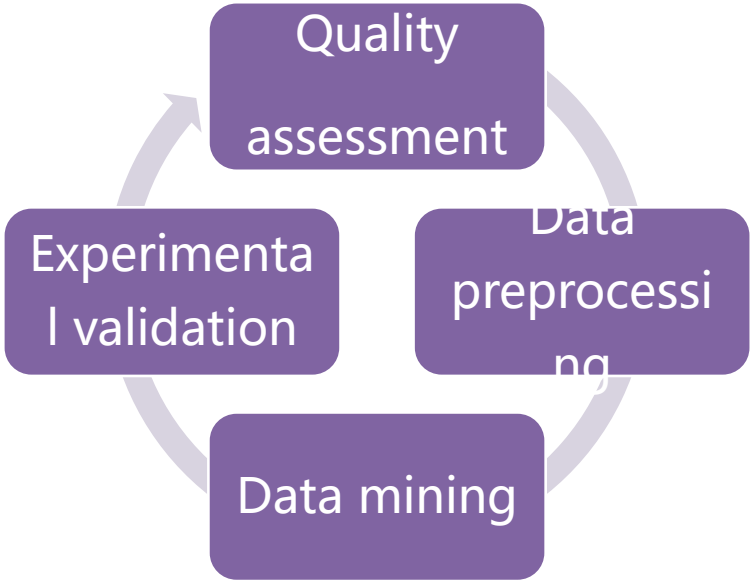
https://en.wikipedia.org/wiki/Scientific_method

Research cycle





Project design → Data analysis → Manuscript preparation → Submission & revision



BGI华大

Things to do in research cycles

- Project design
  - Clarify **scientific question(s)**
  - Choose approaches, i.e. objects + techniques
  - Survey/predict the trend in publication
- Data analysis
  - Data quality control
  - **Repetitive pipeline/code**
  - Plan **experimental validation**

BGI华大 Things to do in research cycles (continued)

- Manuscript preparation
  - Make **self-explanatory figures**
  - Draft **an outline** and follow it
  - Polish the English writing
- Submission & revision
  - Pick a right journal(s), broad or narrow audience
  - Prepare for **major revision**

BGI华大

Abilities required for life science research

- The ability to **read and write scientific papers**
- The ability to **think independently**, and solve problems
- The ability to **employ a variety of techniques**
- The ability to **engage in dialogue** using appropriate scientific language

- Others skills like communication, interpersonal, practical, self-management and so on

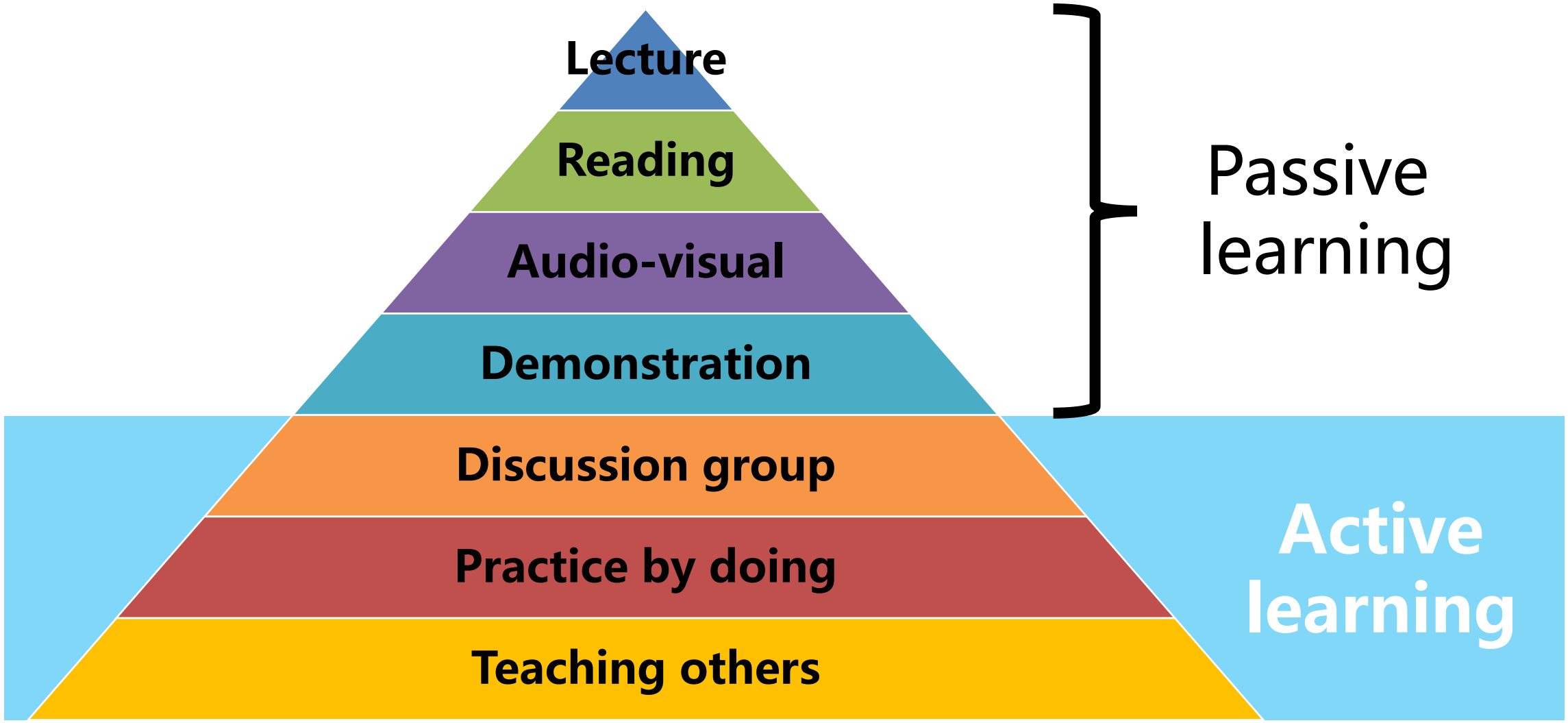
BGI华大

Skills required for research

- **Conceptual/intellectual skills**, read and summarize literature critically, identify and frame **key problems**, think critically and analytically, recognize moral and ethical issues
- **Practical skills**, undertake sufficient work, design/plan/conduct/report investigations, obtain/record/analyze data, conduct work in a safe and responsible manner
- **Communication/presentation skills**, **present research** in a scientific manner, cite other work
- **Interpersonal/teamwork skills**, identify individual and collective goals, debate in a scientific and respectful manner, evaluate individual and team performance
- **Self-management skills**, think and work independently, develop employable skills, develop lifelong learning skills, develop resilience

BGI华大

Learning Pyramid



BGI华大

Education 4.0

Education 1.0 VS Education 4.0

	Education 1.0 Exist in most schools, after-school programs, and private tutoring.	Education 4.0 Used by Dream Formula Education & many leading schools.
Teaching is Done	Lecture-Based (teacher to Student)	Discussion-Based (teacher to student, students to students)
Result	Passive Learning Attitude & Low Engagement	Active Learning Attitude & High Engagement
Teaching Approach	One-size-fit-all	Personalized Education
Result	Low Interest in Learning & Ineffective Learning	High Interest in Learning & Effective Learning
Primary Role of Teachers	Lectures	Mentors
Result	Support Students' Academic Only	Support Students' Holistic Growth
Teacher Training	Degrees, Credentials, Certificates	Continuous Training
Result	Outdated Learning Materials & Teaching Approaches	Stay Updated to Today's Fast-Changing World
Learning Activities	Textbooks & Tests	Project-Based Learning
Result	Unpractical Learning Outcome	Develop Practical 21st-Century Skills
Learning Locations	Traditional Classroom Settings	Large Learning Space
Result	Limited Learning Activities	Diverse Learning Activities
Industry Views Graduates as	Assembly Line Workers	Co-creators & Entrepreneurs
Result	Not Suitable for Today's Industries	Suitable for Today's Industries
Employability	Degree-Qualified & Major-Focused	Prepared for Uncertain Multiple Careers
Result	Not Suitable for Today's Fast-Changing World	Equipped with Variety of Skills for Multiple Career Paths

BGI华大 What professors say about communication

Professor of Statistics, Susan Holmes:

> **Even if you're very, very good with numbers, you also have to understand how to tell your audience what results you obtained. . . It's really important for students to realize that if they want the gift of making the discovery, they have to be the ones who can tell the story.**

Professor of Computer Sciences, Mehran Samani:

> **I've actually seen some methods of extremely powerful computational techniques which, even after they were invented, took years to catch on. Part of the reason why they took that long to catch on was because the people inventing those methods did not do a very good job explaining their methods.**

(Source: Program of Writing and Rhetoric, Stanford University)

BGI华大

What professors say (continued)

- A successful researcher is someone who knows both how to do their science, and how to effectively **communicate** their findings through the **written and spoken** word.

- Effective writing is an important component of a scientist’ s skill set and should be **cultivated at an early stage** of the career.

- This requires a knowledge of the writing and publications processes, and **a mastery of diverse** scientific presentation formats such as research and review manuscripts, abstracts, oral presentations, and social media.

Cell (2022) 47:631-634

Course structure

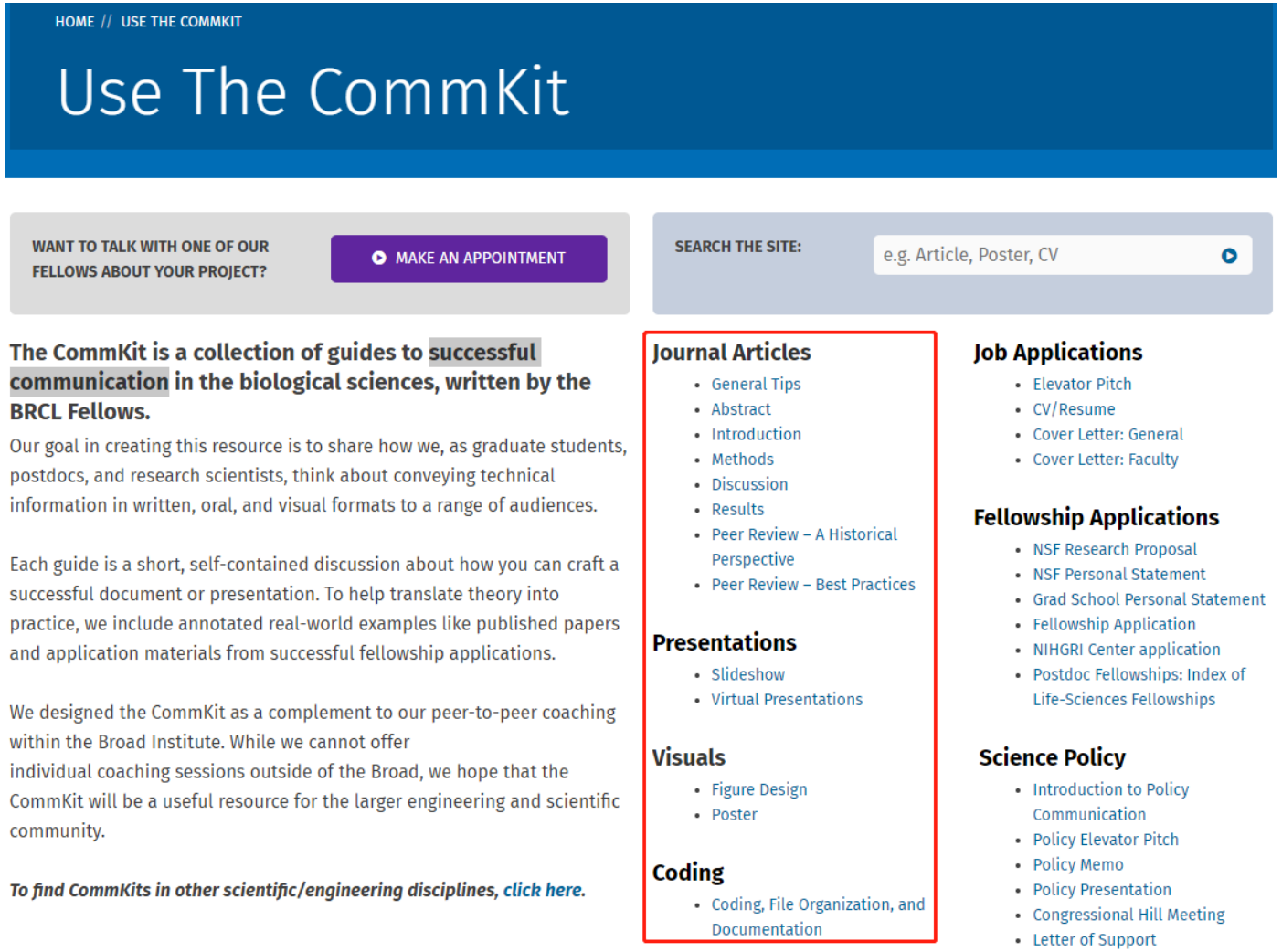
- Instruction
  - Build study groups (~4 students) based on research area, skill, etc.
  - Teach reviewing in the beginning
  - Circulate through groups to answer questions
  - Peer review for small assignments
  - >2 peer review + 2-3 comments from instructors for final
- Timing: spring of the 2nd year
- 2 sessions per week, lecture + workshop
- 2 individual sessions, a short (15-20 min) in the beginning to explore SWOT, and a longer (30 min) in the end to discuss development

Week	Lead instructor	Topic	Assignment	Refs[a]
Week 1	Writing expert	Introduction, Building Confidence	Writing Inventory Describe past writing experiences, strengths, weaknesses (300–400 words)	
Week 2	Writing expert	Writing Science with Style; Editing Sentences	Targeted Editing Edit own past writing and parts of a published paper (Introduction and Discussion)	[i]
Week 3	Writing expert	Titles and Abstracts	Titles and Abstracts Share examples of effective and ineffective titles and abstracts; write conference abstract (200 words)	[1][i–iv]
Week 4	Science expert	Research Proposals Part I: where to submit? selling an effective idea; Specific Aims	Write a Proposal Specific Aims page[b] (1 page, single spaced)	[2][v]
Week 5	Science expert	Research Proposals Part II: significance and innovation; design and methods; peer review and revision	Revise Specific Aims Page Write a peer review of partners aims page (150–200 words)	[2,3]
Week 6	Science expert	Scientific Publication: venues, submission and review process, ethics	Evaluate and Discuss Publishing Venues I: Explore two venues and *BioRxiv* (15 short answer questions) II: Statement regarding preferred publication venue with criteria for choice (250 words)	[4,5][vi]
Week 7	Science expert	Writing a Manuscript Review	Review a Manuscript Select a *BioRxiv* paper to review (maximum 500 words)	[6][vii]
Week 8	Science expert	Research Article Part I: introduction and methods	Introduction Evaluate an introduction (250–500 words)	[1]
Week 9	Science expert	Research Article Part II: results and discussion	Introduction[b] Write an introduction (3–5 paragraphs, 750–1200 words)	[1]
Week 10	Science expert	Data Presentation	Tables and Figures Pick and critique a data figure, table, summary/model figure (300 words)	[7,8][viii,ix]
Week 11	Writing expert	Writing a Review Article	Pitching a Review Article Pitch a review topic for a scientific journal (150 words)	[9][x]
Week 12	Writing expert	Writing for a Lay Audience	Science Tweet Write an effective science-related tweet to a lay audience (280 characters)	[xi–xiv]
Week 13	Writing expert	Presentations	Prepare a 5-min lightning talk (5 slides)	[10][xv–xviii]
Week 14	All	Lightning Presentations	Record/Upload Presentations Students and instructors view and evaluate (rubric + 25 words/presentation)	
Week 15	All	Course Wrap-up	Final Reflective Writing What have you learned? How has your writing changed? (300–500 words)	

Cell (2022) 47:631-634

BGI华大

MIT CommKit



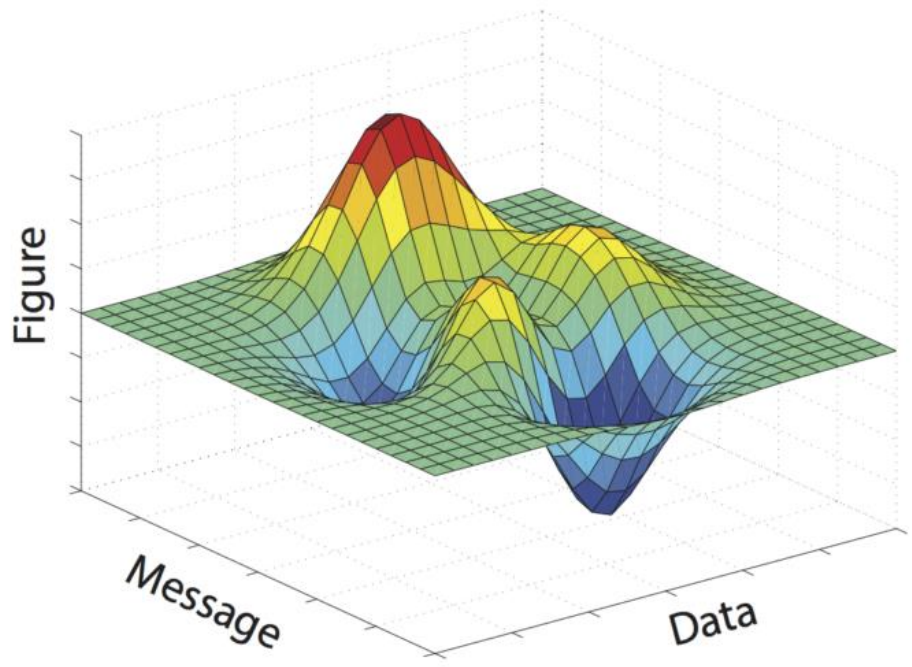
https://mitcommlab.mit.edu/broad/use-the-commkit/

MIT CommKit: figure design

Criteria for Success

1. Your figure leaves the audience with a clear, one-sentence main message
2. You provide evidence that directly supports the main message
3. Any content not related to your main message is removed from the figure

Structure Diagram



Analyze Your Visual Platform

Will you be presenting your figure in an academic paper, a poster presentation, an oral presentation? The final format dictates how your audience will interact with the figure, and how much support or explanation you will be able to provide.

Visual platform	Static or dynamic?	What information goes where?
Paper	Static	• Figure and caption should be sufficient for reader to draw a conclusion. Expert readers judge papers' credibility and impact based on figures alone. • Caption's title should state the message. • Remainder of caption should not contain any interpretation, only high-level description of what was done to obtain data in the figure.
Poster	Static	• You are present and can supplement printed information with spoken explanations. • Precede figure with title that states the message. • A caption is often unnecessary: viewers can easily glance at methods to see how data were obtained. • Larger sizing allows more thorough and direct labeling than is possible for papers. Take advantage of this to make your figure more self-explanatory.
Slides	Dynamic (can be animated)	• Slide title should state the message. • Text should be minimized. • Animations can be used to pace delivery of complex figures. • See **Slideshow Presentation** for more specific skills.

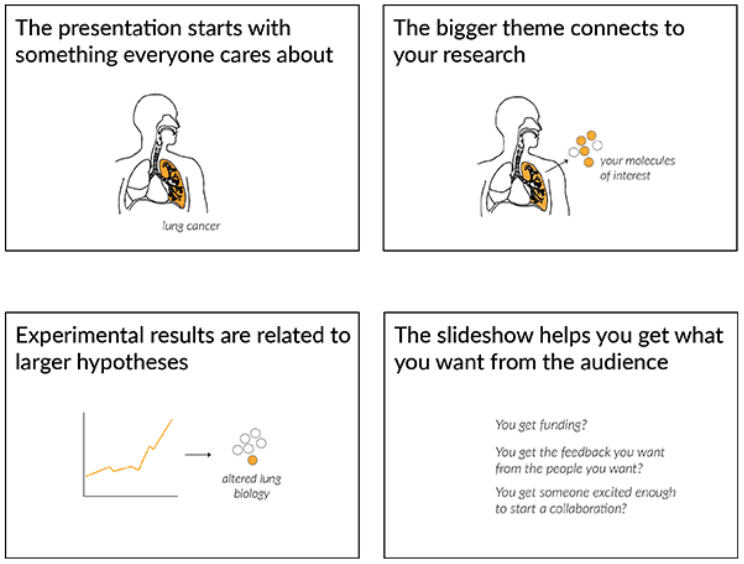
https://mitcommlab.mit.edu/broad/commkit/figure-design/

MIT CommKit: slide show

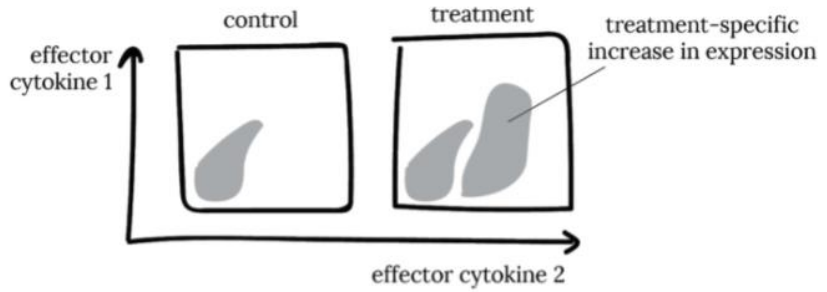
Document Map

1. What is my purpose for giving this presentation?
2. Who is my audience?
3. What are the concepts for success?
    a. Connect your work back to broader motivations and hypotheses
    b. "Introduce" your data
    c. Each slide should convey a single point
    d. Emphasize visuals over text
    e. Make each figure as simple as possible while still conveying its message
    f. Avoid jargon, textual and visual
    g. Prepare for the talking part of the talk

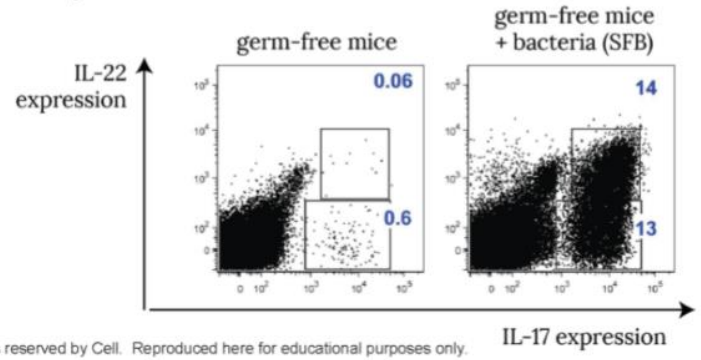
Structure Diagram



Flow cytometry can show if the experimental treatment increases the expression of an effector cytokine



Adding the bacteria of interest increases the expression of effector cytokine IL-17



Try visually explaining unfamiliar data. For an audience unfamiliar with reading FACS plots, an imaginary immunologist first shows a slide explaining how the biological signal would appear in a FACS plot (*top*) before showing the slide with the actual data (*below*). [Adapted from Ivanov *et al.*, *Cell* (2009) doi:10.1016/j.cell.2009.09.033]

https://mitcommlab.mit.edu/broad/commkit/slideshow/

About this course

- The goal is to learn,
  - How to **structure** data and results
  - How to **write** a manuscript (or at least parts of it)
  - How to engage into **team work** on manuscript

- The objectives include,
  - The **methodology** of scientific writing
  - The **habit of periodic summarization** of methods, results, and literature

About this course (continued)

- 45-minute talk by tutor

- **3 reports & 2 presentations** by each student
  - **1-page research proposal** (background, aims, methodology, outcomes, timetable) + **5-minute opening presentation**
  - **3-5-page midterm report** (methods + figures/legends) + **20-minute research presentation**
  - **Final draft** (IMRaD) + **30-minute group sharing**

BGI华大

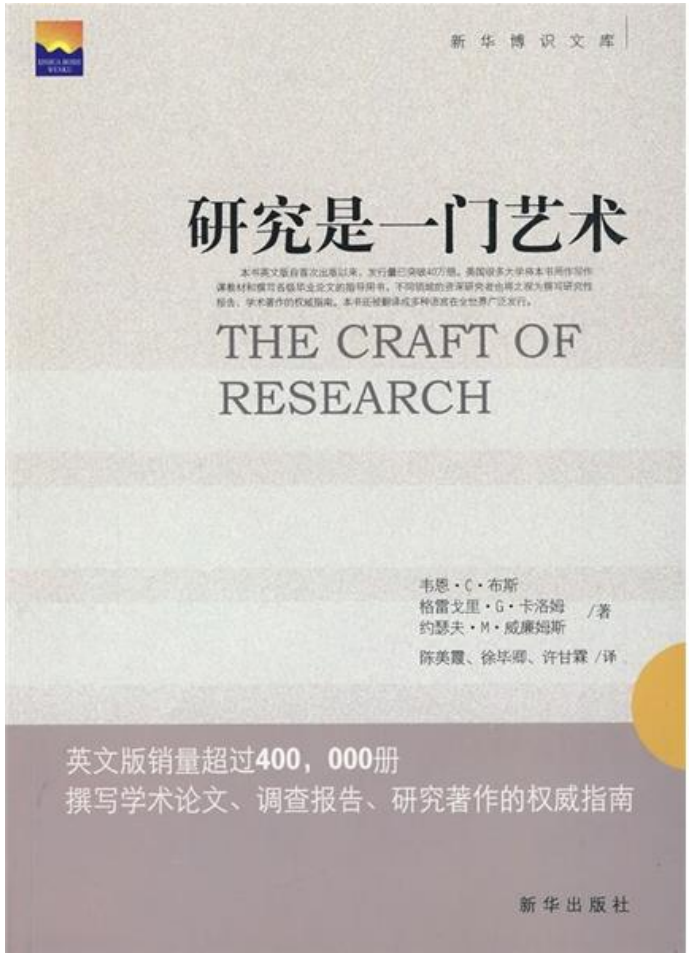
Learning materials: websites

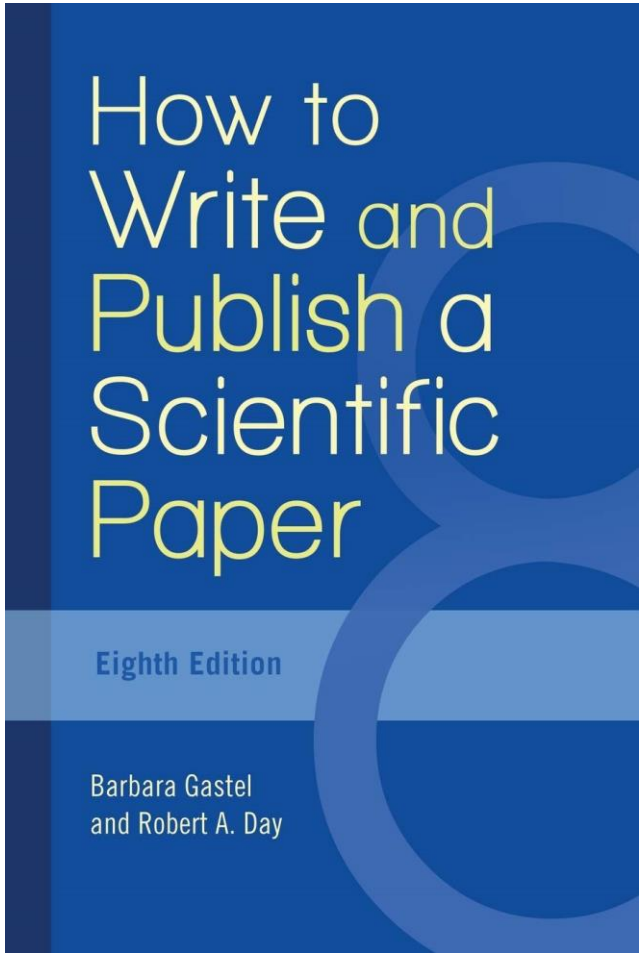
- New paradigm for teaching scientific writing in STEM, Cell (2022) 47:631-634. (https://doi.org/10.1016/j.tibs.2022.03.019)

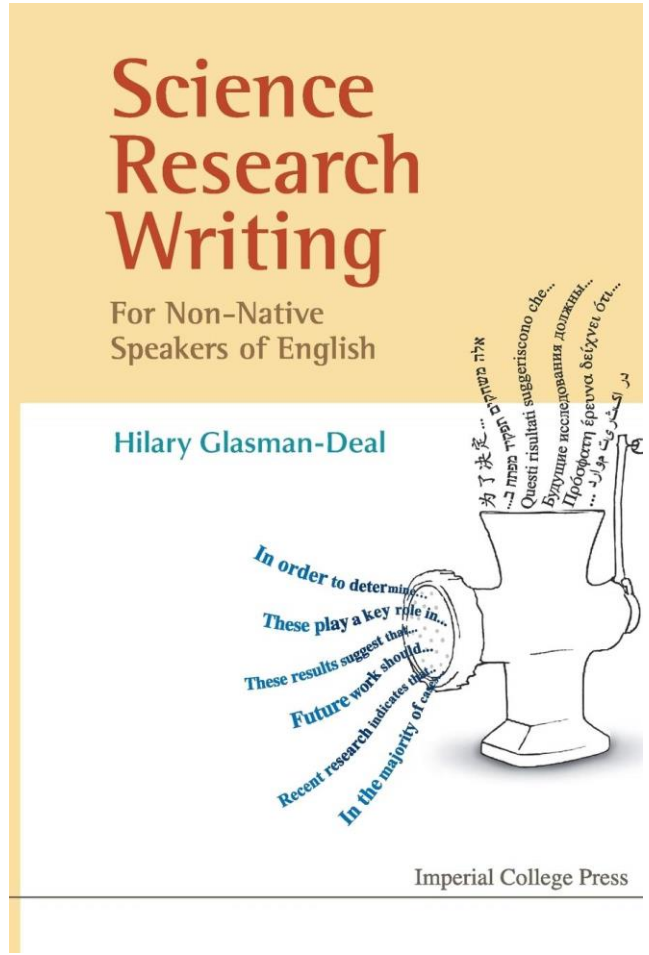
- Stanford Writing in the Sciences, https://www.coursera.org/learn/sciwrite
- MIT CommKits, https://mitcommlab.mit.edu/find-comm-labs/
- UCLA Writing workshop, https://gwc.gsrc.ucla.edu/resources/writing-for-publication
- Northeastern Univ Writing in STEM, https://cssh.northeastern.edu/writingcenter/writing-resources/writing-in-stem/

BGI华大

Learning materials: books







BGI华大

Skills trained in this course

- Intellectual skills
  - Read and summarize literature critically
  - **Identify and frame key problems**
- Practical skills
  - Obtain, record, and analyze data
  - **Present research** in a scientific manner
  - **Draft and revise** a research article
- Interpersonal/teamwork skills
  - **Communicate** in a scientific and respectful manner
  - **Evaluate** individual and team performance

BGI华大

How to thrive in this course

- Propose, **write**, and present a research project
- **Peer review** within each group
- **Peer grading** during course
- **Engage in discussions** during and after course

- Tips: write on your **own work**
- Tips: learn **by practice**, and **from others**

Course website

- About the course
  - https://github.com/popgenome/Writing2022/blob/main/README.md

- Q & A, comment under the following link
  - https://github.com/popgenome/Writing2022/blob/main/Q&A.md

BGI华大

Syllabus

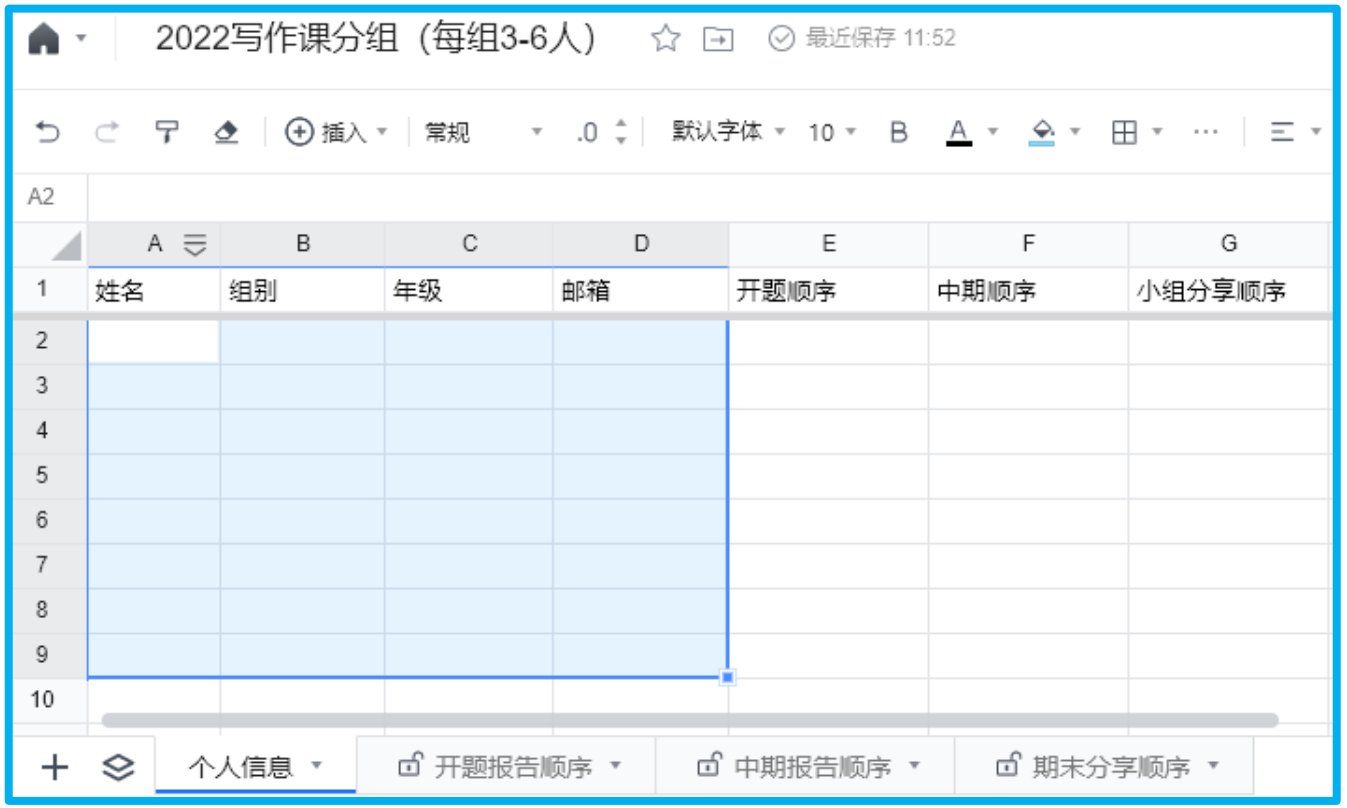
	Course	Assignment(s)
Week 1 (10/17)	Introductory talk + grouping	**1-page research proposal + (optional peer review)** by 10/31
Week 2 (11/2)	**Proposal talk** (10 min) + Scientific figure	**Midterm report w/ peer review** by 11/30
Week 3 (11/9)	**Proposal talk** (10 min) + Group discussion	
Week 4 (11/23)	**Presentation I** (20 min) + Manuscript preparation	**Final report w/ peer review** by 12/31
Week 5 (11/30)	**Presentation II** (20 min) + Submission	
Week 6 (12/21)	**Team sharing** (30 min) + End talk	Student survey

Grading

- Research proposal (20%)
  - **Proposal** (10%), tutor grading based on **format** (5%) + **timing** (5%)
  - **Presentation** (10%), **peer grading** (format 2%, background 4%, methodology 2%, presenting 2%)
- Midterm (30%)
  - **Report w/ edits** (20%), tutor grading based on **format** (5%) + **timing** (5%) + **≥3 peer-review** (10%)
  - **Presentation** (10%), **peer grading** (format 2% + background 2%, results 4%, presenting 2%)
- Final (30%)
  - **Report w/ edits** (20%), tutor grading based on **content** (10%) + **timing** (5%) + **≥3 peer review** (5%)
  - **Team sharing/discussion** (10%), tutor grading based on **attitude**
- Engagement (20%)
  - **Peer review + grading** (1% each)
  - **Q & A** during course + on the github (1-2% each)

BGI华大

To-do: grouping



BGI华大

To-do: presentation booking

时间段	姓名	报告题目
2:00-2:10		
2:10-2:20		
2:20-2:30		
2:30-2:40		
2:40-2:50		
2:50-3:00		
3:00-3:10		
3:10-3:20		
3:20-3:30		
3:30-3:40		
3:40-3:50		
3:50-4:00		
4:00-4:10		
4:10-4:20		
4:20-4:30		
4:30-4:40		
4:40-4:50		
4:50-5:00		

个人信息 | 开题报告顺序

日期	时间段	姓名	报告题目
day1	2:00-2:20		
	2:20-2:40		
	2:40-3:00		
	3:00-3:20		
	3:20-3:40		
	3:40-4:00		
	4:00-4:20		
	4:20-4:40		
	4:40-5:00		
day2	2:00-2:20		
	2:20-2:40		
	2:40-3:00		
	3:00-3:20		
	3:20-3:40		
	3:40-4:00		
	4:00-4:20		
	4:20-4:40		
	4:40-5:00		

个人信息 | 中期报告顺序 | 开题报告顺序

时间段	小组	报告题目
2:00-2:30		
2:30-3:00		
3:00-3:30		
3:30-4:00		
4:00-4:30		

个人信息 | 期末分享顺序

BGI华大

To-do: peer grading

Proposal grading

Midterm grading

BGI华大

Peer grading (continued)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	同学	得分	同学1	同学2	同学3	同学4	同学5	同学6	同学7	同学8	同学9	同学10	同学11	同学12
2		小组	组1	组1	组1	组2	组2	组2	组3	组3	组3	组4	组4	组4
3		同学评分中值												
4		同学1												
5		同学2												
6		同学3												
7		同学4												
8		同学5												
9		同学6												
10		同学7												
11		同学8												
12		同学9												
13		同学10												
14		同学11												
15		同学12												

BGI华大

Final grading

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	姓名	年级	组别	得分汇总	开题报告（20%）		中期报告（30%）		年终报告（30%）		课堂参与（20%）					
2					报告评分	课堂评分	报告评分	课堂评分	报告评分	小组汇报	第一周	第二周	第三周	第四周	第五周	第六周
3																
4																
5																
6																
7																
8																
9																
10																
11																
12																
13																
14																

Part II
Scientific Paper

华大生命科学研究院
BGI · research

BGI华大

Contents in a research paper

- Know **the structure**
  - Title page, Abstract
  - **Main text in IMRaD** : Introduction, Methods & Materials, Results, and Discussion
  - Other materials: supplementary/supporting materials, data and code, peer review information, etc.

- Understand **the logic** in,
  - Introduction
  - Results + Figures/Tables
  - Discussion

Logic in a paper

- **Why** did they do it?
- **How** did they do it?
- **What** did they get?
- **So what** did it mean?

BGI华大

Key questions linked sections

- **Background + question**
- **A sentence or phrases about methods**
- **Major discoveries**
- **Conclusion + significance**

Abstract

WHY → Introduction

HOW → Methods

WHAT → Results

SO WHAT → Discussion

BGI华大

Summary for research articles

- Abstract is the essence of the paper
- Introduction sets the stage
- Methods provides details to repeat the work
- Results describe the data and new findings
- Discussion interpretates in the context

BGI华大

Other article types

- Review
- Technical advances
- Data notes
- Editorial, Perspective, Comments, etc.

- Resources: PUBMED, google scholar, sci-hub

Part III
Research Proposal

华大生命科学研究院
BGI · research

BGI华大

Word template

Title

Name
Affiliation

·Background & Significance

Something Is important, but the mechanism remains unclear. It would bring new insights using some approach.

·Aims

1. To do something
2. To do something else

·Research design

.Time table

.Preliminary data (optional)

.references

题目

姓名
部门

·研究背景与意义

.研究目标

.研究内容

.时间安排

.研究基础（可选）

.参考文献

Revise and evaluate

Proposed Topic/Title of Research:

——Genome assembly of a wild soybean accession from Heil Longjiang Province.

Background:

(Please provide background material that explains the motivation for and aims and objectives of your proposed study, outline the theoretical framework that forms the basis of your research, and specify the research question(s) you intend to investigate.)

Soybean (*Glycine max*), an annual herb, is an important protein and oil source. Historical Previous works evidence indicatesuggesting that cultivated soybean was introduced from north-eastern China approximately *c.* 2510 BP.[1] Compared with domesticated cropscultivars, their wild relative species maintain a higher level of genetic diversity. Since wild species have been challenged in natural environments for thousands of years and maintain a much higher level of genetic diversity, they represent important gene pools, especially for biotic and abiotic stress tolerance, for crop breeding. For example, wild germplasm have been extensively used in soybean breeding programs[2-3]. [2-3]which means north eastern wild soybean is a valuable source for crop breeding.

魏桐(Weitong)
带格式的: 行距 1.5 倍行距

魏桐(Weitong)
设置了格式

魏桐(Weitong)
设置了格式

魏桐(Weitong)
设置了格式: 字体: Arial

魏桐(Weitong)
设置了格式: 字体: Arial

魏桐(Weitong)

BGI华大

PowerPoint template

Title

Name
Affiliation
2022/xx/xx

华大生命科学研究院
BGI · Research

BGI华大

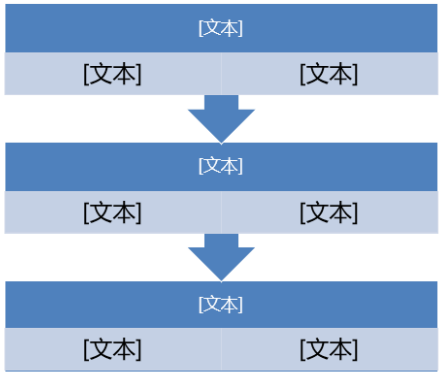
背景介绍

- 科学问题1：问题问题问题问题问题问题?
- 科学问题2：问题问题问题问题问题问题?

BGI华大

课题设计

- 实验设计



BGI华大

时间安排

- 2022/10/xx-xx, proposal writing
- 2022/11/xx-xx, figure preparation
- 2022/11/xx-xx, method/result writing
- 2022/12/xx-xx, discussion/introduction writing
- 2022/12/xx-xx, text polish

BGI华大

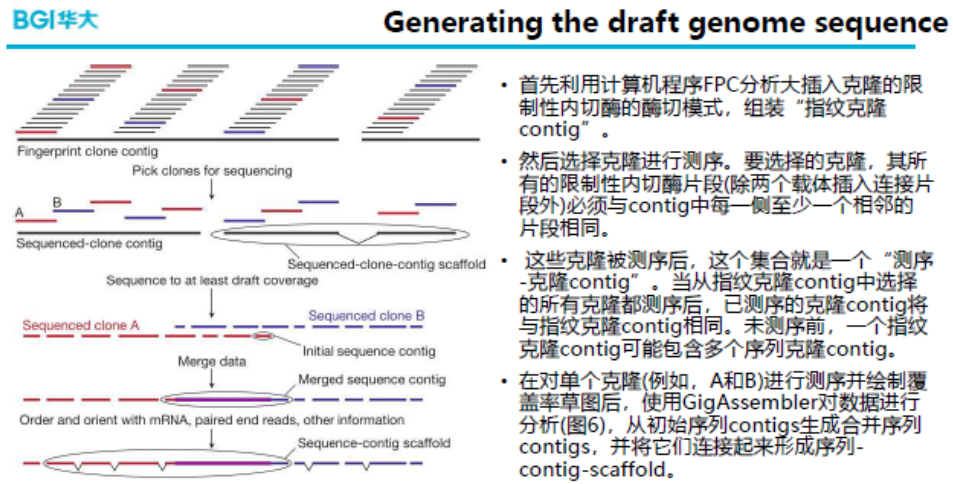
Tips in slides preparation

- Follow **a logical structure**, with focus on major discoveries
- Use **figures and tables** with an appropriate font size
- Use **bullet points** using a big chunk of text
- Keep consistency in slides, font, size, color, etc.

- **Practice, practice, practice…**

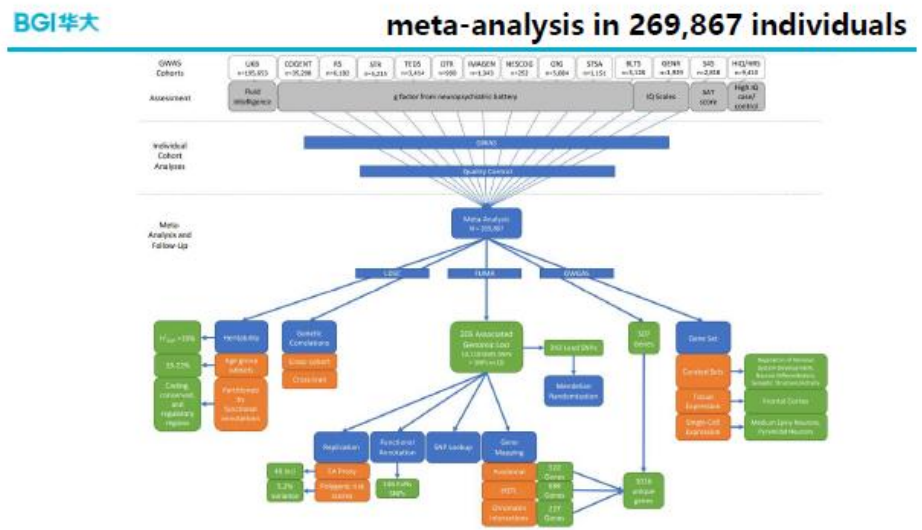
SRaP2021: less is more

1st presentation



- 首先利用计算机程序FPC分析大插入克隆的限制性内切酶的酶切模式，组装“指纹克隆contig”。
- 然后选择克隆进行测序。要选择的克隆，其所有的限制性内切酶片段(除两个载体插入连接片段外)必须与contig中每一侧至少一个相邻的片段相同。
- 这些克隆被测序后，这个集合就是一个“测序-克隆contig”。当从指纹克隆contig中选择的所有克隆都测序后，已测序的克隆contig将与指纹克隆contig相同。未测序前，一个指纹克隆contig可能包含多个序列克隆contig。
- 在对单个克隆(例如，A和B)进行测序并绘制覆盖率草图后，使用GigAssembler对数据进行分析(图6)，从初始序列contigs生成合并序列contigs，并将它们连接起来形成序列-contig-scaffold。

2nd presentation



SRaP2021: one thing at a time

1st presentation

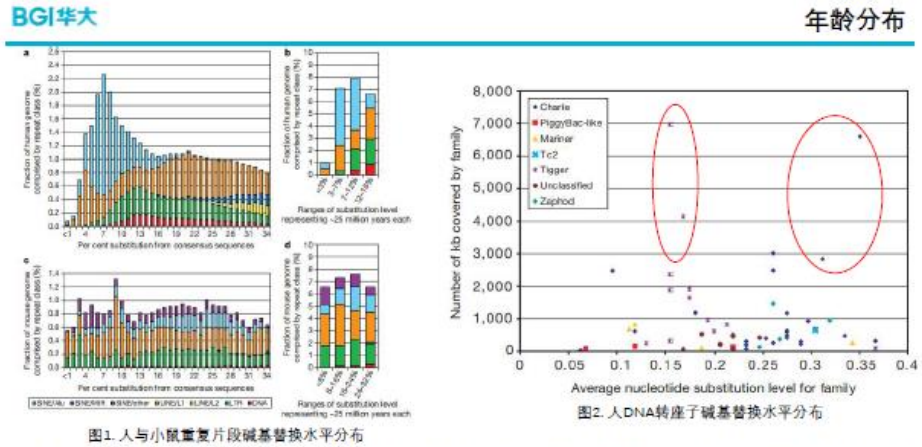
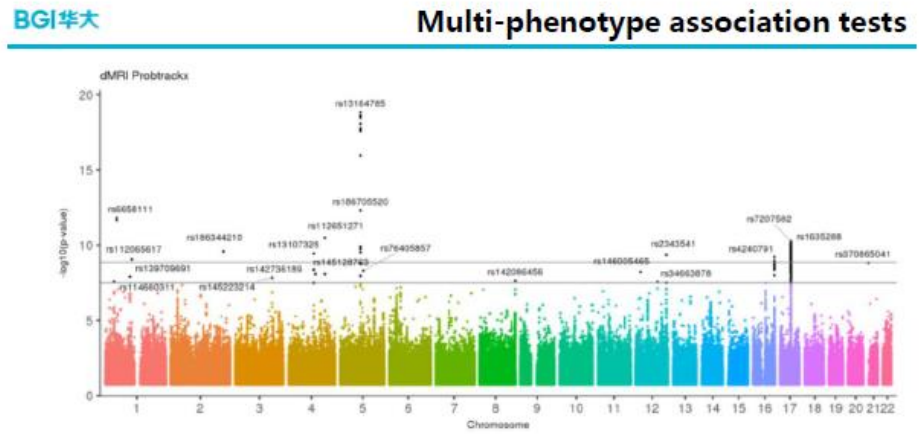


图1. 人与小鼠重复片段碱基替换水平分布

图2. 人DNA转座子碱基替换水平分布

结论：1. 从脊椎动物基因组中清除非功能序列的速度非常慢；2. LINE和SINE的寿命很长；3.DNA转座子有两个峰。由于DNA转座子可以产生大规模的染色体重排，很可能是该活动参与了物种形成事件；4. 序列草图中识别带有功能的全长LTR拷贝只有3个，可能已经快要消失；5.所有转座子的整体活性在过去5000万年的时间里显著下降

2nd presentation



23组IDPs进行多性状GWAS的GWAS曼哈顿图。