Project Title

Establishment of the Regional Multidisciplinary Genomic Epidemiology Center for Research and Training in Infectious Diseases in the Age of Big Data

ชื่อโครงการ

การจัดตั้งศูนย์สหสาขาวิชาด้านระบาดวิทยาในระดับจีโนมเพื่อการศึกษาและวิจัยโรคติดเชื้อด้วยข้อมูลระดับมหสารสนเทศ

ผู้ร่วมโครงการ

มหาวิทยาลัยมหิดล

ประตาป สิงหศิวานนท์ (ที่ปรึกษาโครงการ)

จรณิต แก้วกังวาล

สารนาถ ล้อพูลศรี นิยม

อำนาจ คำศิริวัชรา

ธนรรถ ชูขจร (ผู้ประสานงานโครงการ)

Big Data Institute, Oxford University

Dominic Kwiatkowski

Olivo Miotto

รายละเอียดการติดต่อ

หน่วยจีโนมและเวชศาสตร์วิวัฒนาการ ศูนย์การวิจัยมาลาเรีย คณะเวชศาสตร์เขตร้อน มหาวิทยาลัยมหิดล

โทร. 02-354-4086, 085-155-2466 E-mail. thanat.cho@mahidol.edu

Summary

Our overall vision is to develop the regional center for the research and educational paradigm on the use of Big Data to combat infectious diseases. With the rise of the large-scale data from thousands of pathogen genomes, the new generation of researchers must be able to mine the genetic data and monitor microbe evolution in the population scale. This new set of skills will allow them to detect genetic changes in real-time, leading to the capability to thwart potential threats before it even happens.

In this proposal, we aim to develop this capacity by creating multidisciplinary research and training center. Under the support of the Bill and Melinda Gates Foundation (BMGF), members of the Big Data institute and the Wellcome Sanger Institute will work with the Genomics and Evolutionary Medicine Unit at Mahidol University to establish a regional genomic surveillance center and train a local coding team. The research focus of this team is to decipher genomic information of infectious diseases common in Southeast Asia and translate the population-scale genome sequences into actionable data for the national governmental agencies. The Gates Foundation aims to develops the local capacity to use the genomic epidemiology data in malaria elimination. Nevertheless, this capacity can be applied to the control of circulating pathogens and emerging infectious diseases in Southeast Asia. We will also train Southeast Asian researchers to gain hand-on experience working with the genomic data from drug-resistant parasites collected from Southeast Asia with the Thai public health community and malaria network in neighboring countries. The new graduate program track on genomic epidemiology will be developed to train the partners from Cambodia, Laos, Myanmar and Vietnam. Our trainees will become the new-generation of Thai public health officials and researchers with the expertise in Genomic Epidemiology. The trainees will be parts of the strong and ongoing collaborations between the Faculty of Tropical Medicine (TropMed) and key governmental agencies, including the Ministry of Public Health and the Ministry of Agriculture and Cooperatives. The trainees will learn how to apply large-scale genomic epidemiology data into these collaborations. The advanced training courses and direct hand-on research experience will generate a group of skilled genome epidemiologists and genome scientists for Mahidol University and Thailand.

This regional center in genomic epidemiology is a part of the growing collaborations between TropMed, the Big Data Institute at Oxford University and the Wellcome Sanger Institute. During the last three years, members of the Institutes and TropMed jointly organized international training courses and co-supervised graduate students and postdoctoral fellows. TropMed and the Big Data Institute will strengthen our training effort and develop a sustainable training structure as described in the proposal. The support from Mahidol University will be used to provide laboratory setting for genomic analysis, and the support from the partner will be in the forms of sequencing machines, reagents and technology knowhow. The project detail with measurable output and outcome will be described in the proposal. The support will be critical for making Mahidol University the center of Big Data Genomic Epidemiology in Asia.

1. Establishing the Research Network and Regional Center

The cooperation between the members of TropMed and the Big Data Institute at Oxford University will introduce the big data and genomic elements into the research and educational paradigm at Mahidol University. The proposal is built from the fruitful collaborative activities during the last few years between the researchers and students of the two institutes. For example, Drs. Dominic Kwiankowski and Olivo Mitto, members of the Big Data Institute, have been playing an active role in research supervision of the students in Dr. Thanat Chookajorn's Unit at TropMed and have already generated two Ph.D. graduates. At present, the team is now supervising three trainees at the graduate and postdoctoral levels. With our previous accomplishments, we would like to expand the training and research opportunities to the Thai scientific community. The activity of our network will be described as follows:

Investment in building the Regional Center in Genomic Epidemiology

At present, the Gates Foundation is promoting the use of the genomic epidemiology data in malaria elimination. The goal is for each endemic country to establish a local team and to develop network for data sharing. The support for building a local team for both genomic sequencing and computer coding is a part of the plan to for developing a regional cooperation. The package will include the equipment for Illumina-based genomic sequencing and reagents (approximately 6 million baht in the first year). In addition, the cost will include the salary support for two advanced trainees and for their training abroad as well as the training course held in collaboration with TropMed (approximately 1 million baht per year). The support from Mahidol University is to provide the fund to match the support from BMGF. At present, BMGF plans to invest in the analysis project under the collaboration between the Big Data Institute, the Wellcome Sanger Institute and Mahidol University. There is an initiative to develop a local capacity for whole-genome sequencing capability. The support here from Mahidol University will show our partner that the sequencing facility with the technical support for genomic epidemiology surveillance should be based at Mahidol University. The support from Mahidol University will include the set up for genomic surveillance research which will make the site attractive for the investment from BMGF and every partner.

Practical and academic courses

The Big Data analysis of genomic epidemiology is a multidisciplinary field, consisting of statistics, genome biology, epidemiology, population genetics and molecular biology. There is no one-size-fits-all approach for the training. We aim to provide the structure to systemically introduce the working knowledge to the trainees from respective fields. We will implement three short courses during the period of twelve months including genomic epidemiology and big data analysis (Table 1). The courses will introduce working principles and have hand-on exercises to allow trainees to be exposed to real research problems. The courses will be opened to the staffers and students at Mahidol University. We will simultaneously develop a long-term training

structure to allow trainees to sustainably develop the skills. We will work with other Faculty members to introduce genomic epidemiology into existing courses on epidemiology and bioinformatics. We contact two course coordinators on implementing the lectures into the graduate study courses at TropMed. We will use the information and feedback collected from the courses to develop a new training track for the graduate programme at TropMed. The courses will be joined by distinguished members of the Big Data Institute as guest lecturers and co-supervisors.

Training the trainers with hand-on experience and public engagement

It is undeniable that, to generate a long-lasting impact, we need to build a new generation of experts that can share the knowledge with the community. We realize that, to become an expert on new multidisciplinary field, trainees require direct hand-on experience in the nurturing environment. We will accept two trainees to work on a collaborative project between TropMed and the Big Data Institute. One of our research goals is to convert the large-scale genomic data collected from thousands of malaria parasites throughout Southeast Asia into a simple report card for the national malaria control program. The report card will be generated by using a decision-making algorithm to decipher genomic data into drug susceptibility profiles. We will accept the two Thai trainees at the advanced level to take part in this project. We will also provide four internship positions for four undergraduate trainees in the first year to help in the project. An advanced course for University lecturers and advanced school teachers will also be arranged to introduce the educational materials on how to teach genomic and big data analysis in the form of virtual box that can be applied at the local schools and respective Faculties.

2. Continuity and Impact

The support from the University will provide the seed to establish a formal collaboration toward forming a center dedicated to the Big Data analysis of the genomic information with direct impact to the medical community. The establishment of the regional center and the training structures will play a critical role toward the goals. We envision that the skill and cooperation will be expanded to both communicable and incommunicable diseases that affect humans in the population scale. Our plan is developed to ensure that the output and the outcome will have direct impact on the Thai public health community. At present, there is an ongoing programme to record the malaria cases from every province in Thailand. The database is developed by members of TropMed. The genotype data can be integrated into the workflow in order to guarantee the long-term sustainable contribution of this project. We will also work toward developing a new training track for the Big Data application in infectious diseases. The materials generated during the training period will be packaged into the academic course structure.

The hand-on research project for the advanced trainees will base on genomic surveillance of malaria drug resistance. The status quo of malaria treatment protocols is to nationally employ a combination of drugs specific to parasite species. The usefulness of the national drug regimen has often run its course by the eventual rise of drug-resistant parasites. Thailand has successfully implemented a nation-wide antimalarial regimen in order to ensure a coherent antimalarial control policy. The government provides an incentive in the form of free drugs distributed at governmental health centers. It is enforced by a strict antimalarial import and the law prohibiting the sale by private entities. The policy has been proven to be successful in promoting a combination therapy. Still, a choice of drug combinations is rigid and is historically switched to new antimalarials only when the old one became ineffective.

An alternative to this status quo is that the drug regimens can be adjusted in order to avoid losing the precious antimalarials to multidrug-resistant strains. With the comprehensive genomic data in the population scale and constant resistance surveillance, the effectiveness of the regimens can now be evaluated in real-time. The information can lead to the replacement of a drug regimen to avoid full blown resistance. Switching between regimens could also save precious antimalarials for future use. However, the possibility of adopting a new paradigm is challenged by the point of view that a national drug policy cannot be changed quickly enough to accommodate new scientific data. This exchange of opinions is not merely academics, but it is a real and ongoing discussion at the policy-maker level.

The networking grant brings a unique opportunity to address this issue. The network will be established by bringing public health officials, biologists, epidemiologists and mathematical modelers to work together. Genetic variants associated with changes in antimalarial susceptibility have been identified since 1980s. These identifications lead to useful molecular markers, but the precious drugs are already lost. The technology and the comprehensive clinical surveillance program in Southeast Asia during the last decade provide two unprecedented opportunities. First, due to careful drug combinations and active surveillance, clinicians and scientists can observe the populations of malaria parasites that start losing sensitivity to

antimalarials including but not limited to artemisinin, piperaquine and mefloquine. The current clinical dosage is still recommended, but treatment failure and reduced drug susceptibility profiles are observed. The second opportunity is the comprehensive population data based on whole-genome sequencing of malaria parasites in Southeast Asia. The genetic variations in the forms of SNP and CNV are associated with drug susceptibility reduction. The role of these genetic changes have become more complicated, going beyond conventional causal mutations. The mutations can also act as a compensatory mechanism to improve fitness of drug-resistant parasites. The advanced trainees at the graduate and postdoctoral level will have the opportunity to access the large-scale genomic data and develop the script to translate the data into a simple report card. Four interns will also be recruited to help running the workflow. The trainees will involve with every step of the workflow in order to develop local capacity for Big data research.

The ultimate goal of this plan is to develop the Southeast Asian center for Genomic Surveillance. At present, the Gates Foundation has expressed a strong interest to use genomic data to guide the national malaria control programmes in malaria elimination. The effort will require the implementation by well-trained staffers, and it is inevitable that the regional training structure is required. The course and training structure developed by this proposal will be presented to the foundation as a local contribution to establish a center in Thailand.

Table 1. Instructors and tentative training topics.

TropMed, Mahidol University	
Amnat Khamsiriwatchara	- Data management and visualization in the clinical context
Jaranit Kaewkungwal	- Statistical modeling in Epidemiology
Saranath Lawpoolsri	- Epidemiological study design
Thanat Chookajorn	- Tracking evolutionary changes at the population scale.
Theerarat Kochakarn	- Selection of matching sequencing technology
Big Data Institute, Oxford University	
Dominic Kwiatkowski	- Real-team genomic epidemiology
Olivo Miotto	- Big data storage and analytics for genomic research
TBD	- Data visualization in the Big Data context

4. Output and Outcome

The aforementioned detail is summarized in Table 2.

Table 2. Expected output and outcome.

Period	Activity	Output	Outcome
6 Months	1.Organize training courses	-2 practical courses - 30 trainees	- satisfactory result (>4.2 from 5 scale) - 70% return customers - feedback for revising the material for developing academic courses
	2. Set up workstations for high-throughput sequencing and analysis	- 2 stations for MinION	- Local working capability for high-throughput sequencing
	3. Hand-on training	- two researchers joining the collaborative projects between the Big Data Institute and TropMed	- Two trainees with international publications on big data research
12 Months	4. International training course during the Joint Tropical Medicine Meeting.	- 20 trainees	- satisfactory result (>4.2 from 5 scale) - feedback for revising the material for developing academic courses
	5. Establish a local sequencing facility to be used as a regional surveillance center	- Establish whole-genome sequencing facility	- Genomic analysis of 2,000 malaria samples in the first year
	6. A practical course for the whole semester for the trainees to have hand-on experience from bench to population analysis	- 10 trainees (advanced level)	- satisfactory result (>4.2 from 5 scale) - feedback for revising the material for developing academic courses
	7. A 'training the trainee' course to share teaching materials with the Thai academic community	- 20 trainees	- satisfactory result (>4.2 from 5 scale) - 70% implementation from the survey for material implementation at trainees' institutes
	8. Training researchers in the joint research projects	- two researchers joining the collaborative projects between the Big Data Institute and TropMed	- Two trainees with international publications on big data research - Integrate the big data output into the national malaria surveillance database

4. Collaboration Structure

The collaborators from the Big Data Institute at the University of Oxford will be led by Professors Dominic Kwiatkowski and Olivo Miotto. They are the world-class authority on genomic epidemiology who made several seminal discoveries in infectious diseases. They will be joined by an expert on big data visualization whom we will invite at the latter date. Professor Kwiatkowski also leads the infectious disease programme at Wellcome Sanger Institute, widely recognized as the best genome center in the world. Professor Miotto plays an active role in malaria genomic surveillance throughout Southeast Asia. He is the pioneer in genomic epidemiology of malaria parasites in Southeast Asia. They will work with Assistant Professor Thanat Chookajorn and the members of the GEM unit to implement the plan. The team will also work with Associate Professor Saranath Lawpoolsri and Assistant Dean Amnat Khamsiriwatchara who has generated several disease database for many Thai governmental agencies.

Dean Pratap Singhasivanon and Associate Professor Jaranit Kaewkangwal will be two senior advisors for this project. Dean Pratap is a Thai leading epidemiologist and a physician with expertise in Tropical Medicine. He is the Dean of the Faculty of Tropical Medicine at Mahidol University (TropMed) in Thailand and serves as the Secretary General of SEAMEO Tropmed Network which is a network for training and research in tropical medicine and public health under the Southeast Asian Ministers of Education. He is working very closely with Associate Professor Jaranit Kaewkangwal, an expert in Biostatistics and Epidemiology, at TropMed. They have been working with MOPH to establish several surveillance networks for key infectious diseases in Thailand. Their malaria database is the centralized database for the whole country with a complete record from multiple public health entities. The participation of Professors Singhasivanon and Kaewkangwal in this network will bring experienced hands on how to cooperate with the Ministry. They also have proven track records on the implementation of epidemiology and modeling data to influence national public health policies. The expertise of the UK team on genomic analysis is not available in Thailand and is complementary to their epidemiology and modeling research.

5. Budget

รายการ	จำนวนเงิน	รวม
1. **งบบุคลากร**		
2. **งบดำเนินงาน**		
2.1 **ค่าตอบแทน ใช้สอย และวัสดุ**		50,000.00
2.1.1 ค่าจัดทำบัญชีและรายงาน	50,000.00	
2.2 **ค่าสาธารณูปโภค**		
3. **งบลงทุน** (ระบุรายการ)		1,932,500.00
3.1 Mic qPCR Cycler	963,000.00	
3.2 Denovic QFX Fluorometer	144,450.00	
3.3 BioShake Microplate Shaker (specific to the Illumina protocol)	160,500.00	
3.4 Model 1580R Microplate Refrigerated Centrifuge	535,000.00	
3.5 Magnetic Stand	69,550.00	
3.6 ค่าบำรุงรักษาและควบคุมเครื่องมือ	60,000.00	
งบประมาณที่ขอเสนอ	1,982,500.00	1,982,500.00

หมายเหตุ: ทาง Big Data Institute และ Wellcome Sanger Institute จะร่วมสนับสนุนงบประมาณในการฝึกบุคลากร ประมาณปีละ 1 ล้านบาท และจะประสานงานกับ Bill and Melinda Gates Foundation ในการลงทุนเรื่องเครื่องมือและการตรวจเป็นค่าใช้จ่ายประมาณ 6 ล้านบาทในปีแรก