

# Rising Strengths Hong Kong SAR in Bioinformatics

Chiranjib Chakraborty<sup>1,3</sup> · C. George Priya Doss<sup>2</sup> · Hailong Zhu<sup>3</sup> · Govindasamy Agoramoorthy<sup>4</sup>

Received: 13 September 2015 / Revised: 7 December 2015 / Accepted: 8 January 2016  
© International Association of Scientists in the Interdisciplinary Areas and Springer-Verlag Berlin Heidelberg 2016

**Abstract** Hong Kong's bioinformatics sector is attaining new heights in combination with its economic boom and the predominance of the working-age group in its population. Factors such as a knowledge-based and free-market economy have contributed towards a prominent position on the world map of bioinformatics. In this review, we have considered the educational measures, landmark research activities and the achievements of bioinformatics companies and the role of the Hong Kong government in the establishment of bioinformatics as strength. However, several hurdles remain. New government policies will assist computational biologists to overcome these hurdles and further raise the profile of the field. There is a high expectation that bioinformatics in Hong Kong will be a promising area for the next generation.

**Keywords** Bioinformatics research · Bioinformatics education · Government role · Hong Kong SAR

## 1 Introduction

Concerning Hong Kong's economic development, Tony Fu-Lai Yu, an economics scholar from the University of New South Wales, Australia, once wrote: 'From barren island, Hong Kong changed into the mart of East Asia'. But what puzzled economists is that Hong Kong is only a small region with approximately 6 million people living in an area of around 1064 km<sup>2</sup> [1]. Today, the population of this city has grown to more than 7 million [2]. Hong Kong, a special administrative region (SAR) of the People's Republic of China (PRC), is situated in the southern part of China that is enclosed by the [http://en.wikipedia.org/wiki/Pearl\\_River\\_Delta](http://en.wikipedia.org/wiki/Pearl_River_Delta) and the South China Sea ([http://www.censtatd.gov.hk/FileManager/EN/Content\\_810/geog.pdf](http://www.censtatd.gov.hk/FileManager/EN/Content_810/geog.pdf)). Hong Kong is one of the world's leading financial and trading centres and is also well known as a free trade zone with low taxation.

The area became controlled by the Chinese in 221 BC, during the Qin dynasty of China [3], and remained under Chinese control until after the Anglo-Chinese War of 1839–1842 (also known as the First Opium War) when it became a protectorate of the UK. The territory of Hong Kong was initially restricted to Hong Kong Island, with the British then extending their boundaries in two phases. The first phase, in 1860, saw the British outpost being extended to the Kowloon Peninsula, and in 1898 it was again extended to include the New Territories. Hong Kong underwent its first peaceful and democratic transfer of power in 1997 when China regained sovereignty [4, 5]. The area had been occupied by the Japanese in 1941 after the Battle of Hong Kong, which occupancy ended after 3 years and 8 months [6] when Japan surrendered at the end of the Second World War, in 1945.

---

✉ Hailong Zhu  
hlzhu@comp.hkbu.edu.hk

✉ Govindasamy Agoramoorthy  
agoram@tajen.edu.tw

<sup>1</sup> Department of Bio-informatics, School of Computer and Information Sciences, Galgotias University, Greater Noida, UP 201306, India

<sup>2</sup> Medical Biotechnology Division, School of BioSciences and Technology, VIT University, Vellore, TN 632014, India

<sup>3</sup> Department of Computer Sciences, Hong Kong Baptist University, Kowloon Tong, Hong Kong

<sup>4</sup> College of Pharmacy and Health Care, Tajen University, Yanpu, Pingtung 907, Taiwan

Hong Kong has now become a major global financial centre and one of East Asia's 'Economic Tigers'. The territory is economically and technically in an excellent position to assist developing nations worldwide [7]. During the past two decades, the Hong Kong SAR government has been actively pursuing growth in high-tech industries such as telecommunications, biotechnology, information technology and electronics to help develop this economic leader of Asia into a global technological giant in today's high-tech world.

In Hong Kong, computational biology has adopted bioinformatics as a research and development (R&D) area in its own right. Researchers in the life sciences, biochemistry, molecular biology and computer science have all used bioinformatics as a tool to solve their particular research problems. In this paper, we discuss how the economic growth of Hong Kong and its population age structure have helped to establish computational biology in the territory and have helped with the location of its foundations. To this end, we identify the learning activities, landmark research, activities of bioinformatics companies and role of the government in Hong Kong in establishing computational biology as strength. We have tried to be comprehensive and apologise to any researcher or research group that has not been included in this report.

## 2 Hong Kong's Economic Boom and Population Age Structure Assist the Establishment of Computational Biology

Hong Kong has a free-market economy which is highly dependent on international trade and finance. The USA is Hong Kong's second-largest export partner after mainland China, which is an important factor in the Hong Kong economy ([http://www.tid.gov.hk/english/trade\\_relations/mainland/trade.html](http://www.tid.gov.hk/english/trade_relations/mainland/trade.html)). During the past decade, the GDP of Hong Kong has grown steadily and is now about 6.0 % (Fig. 1). The territory has one of the highest per capita incomes in the world. This dynamic economic environment provides an added advantage in high-tech R&D areas such as computer science and computational biology [8, 9].

The economy of any nation depends on its age structure, which presents various challenges. The age structure of a population is directly related to that society's productivity and economy [10]. Hong Kong has an advantage in that the percentage of its population that is of working age is one of the highest in the world. Hong Kong's working-age population (15–64 years) is about 74.8 %, and the median age in the territory is 43.4 years. The older age group (65 years and over) comprises about 13.5 % of Hong Kong's population (Fig. 1). It has been observed that a higher

proportion of older people leads to a critical shortage in manpower and tends to undermine productivity. The healthcare costs of an older population are noted to be economically unproductive, which is also a burden for society [10].

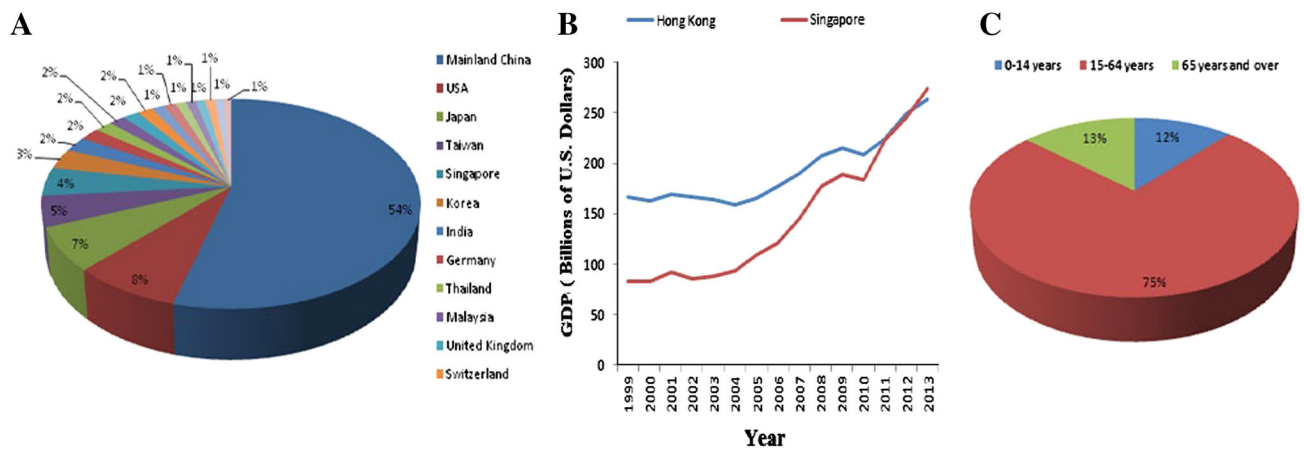
Factors such as these highly competitive business environments, the knowledge-based economy, the best-performing economy in Asia in terms of investment and the vibrancy of the working-age group assist in Hong Kong's rapid growth to become a strong economic force in East Asia and to emerge as a global technological giant able to emphasise its pioneering success in computational biology.

## 3 Foundations of Bioinformatics in Hong Kong

The initial step in establishing bioinformatics was the commencement of the Hong Kong government's innovation and technology development programme in 1998–1999, which provided funding of about \$5 billion through the Innovation and Technology Fund (ITF). Four types of project are financed through the ITF programmes of innovation and technology support, university–industry collaboration, general support and small entrepreneur research assistance, all involving innovation or technology that should modernise the industry (<http://www.gov.hk/en/about/abouthk/factsheets/docs/technology.pdf>).

In the late 1980s, especially 1988, work on bioinformatics in Hong Kong was begun by a group of researchers in the life sciences. Two projects are of particular note: first, the use of microcomputers in histopathology [11] and second, the development of the commercial Microbact 24E (MB24E) microsystem for the identification of common clinical isolates of *Enterobacteriaceae* and non-fermenting Gram-negative bacilli [12]. In 1998, two more bioinformatics papers by different research groups were published in the journal *Bioinformatics*. Smith and Xue, from the Biochemistry Department of the Hong Kong University of Science and Technology (HKUST), developed a method for summing up and presenting the information contained in a set of aligned sequences to identify patterns within the sequences and represent them in a more accurate and graphical form [13]. Chau et al. [14], from the Department of Applied Biology and Chemical Technology at the Hong Kong Polytechnic University (PolyU), developed a software package entitled 'TLCQA' for low-cost analysis of thin-layer chromatography images. Computational biology had thus already spread to the different disciplines within Hong Kong.

The Hong Kong Bioinformatics Centre, jointly established by the Chinese University of Hong Kong (CUHK) and the HKUST in 1998 and funded by the Innovation and Technology Commission (ITC) of Hong Kong, was the first



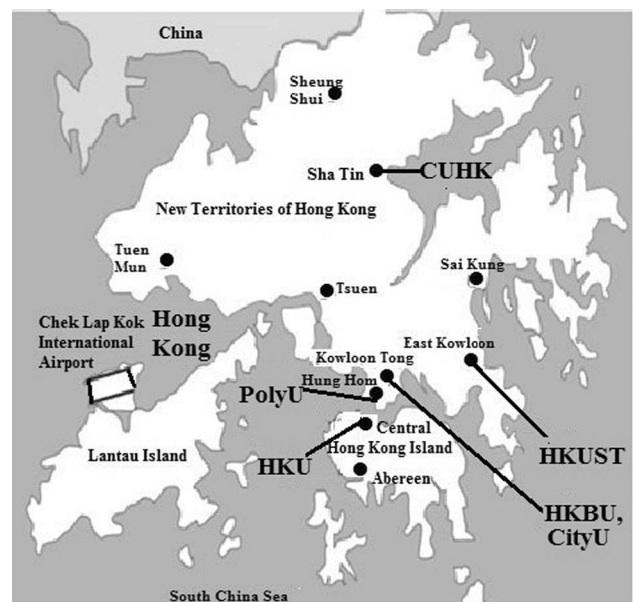
**Fig. 1** Hong Kong's robust economy and its population age structure. **a** Hong Kong exports by percentage to different partner countries (data source: <http://www.tid.gov.hk/>). **b** Comparison of

GDP between Hong Kong and Singapore (data source: <http://www.tradingeconomics.com>). **c** Age structure of the population in Hong Kong

bioinformatics research centre in the Hong Kong SAR. This research centre acts as a central source of information to support the biotechnology community in this region; its current director is Professor Tsui Kwok Wing Stephen (<http://www.hkbic.cuhk.edu.hk/>).

#### 4 Bioinformatics Education, Workshops and Conferences in Hong Kong

Universities in the Hong Kong SAR, such as the University of Hong Kong (HKU) ([www.hku.hk](http://www.hku.hk)), the CUHK ([www.cuhk.edu.hk](http://www.cuhk.edu.hk)), the HKUST ([www.ust.hk](http://www.ust.hk)), the Hong Kong Baptist University (HKBU) (<http://www.hkbu.edu.hk/>), the City University of Hong Kong (CityU) (<http://www.cityu.edu.hk/>) and the PolyU ([www.polyu.edu.hk](http://www.polyu.edu.hk)), assist in providing manpower training and development in the field of computational biology (Fig. 2). Currently, several undergraduate or postgraduate programmes or courses are available in the specific area of bioinformatics (Table 1). The departments of Biochemistry and Computer Science at HKU offered the first undergraduate programme, a B.Sc. in Bioinformatics, in September 2001; this combined programme of biochemistry and computer science was supported by both departments. This course is not currently being offered by HKU. Instead, HKU's Department of Biochemistry offers 'Sequence Bioinformatics' (course code: BIOC3608), a course for B.Sc. students that carries 6 credit points. This course was planned to enable students to gain a full knowledge of sequence bioinformatics. Outcomes of the course include knowledge of algorithms for sequence alignments, BLAST searches and construction of phylogenetic trees. A course with the same name but a different course code (BIOC3605) is offered by the Faculty of Medicine at HKU for a Bachelor Degree in Medical Science. Although



**Fig. 2** Location of the major universities in Hong Kong which provide manpower training in computational biology. HKU University of Hong Kong, CUHK Chinese University of Hong Kong, HKUST Hong Kong University of Science and Technology, HKBU Hong Kong Baptist University, CityU City University of Hong Kong, PolyU Hong Kong Polytechnic University

this course also carries 6 credit points, the course syllabus is different; students learn about computational methods for analysis of DNA and protein sequences. In addition, the Computer Science Department in the Faculty of Engineering at HKU offers a course entitled 'Computational Molecular Biology' (course code CSIS0326), also carrying 6 credit points, for students pursuing a Bachelor of Engineering degree. The course teaches the students how to use algorithms to work out computational problems related to the enormous

**Table 1** Examples of bioinformatics training courses, graduate programmes and other bioinformatics courses in Hong Kong

University	Website	Department/School	Programme/Course	Remark
University of Hong Kong (HKU)	<a href="http://www.hku.hk">www.hku.hk</a>	Department of Biochemistry	Course 'Sequence Bioinformatics' (course code BIOC3608)	6 Credit point course for B.Sc. students
		Faculty of Medicine	Course 'Sequence Bioinformatics' (course code BIOC3608)	6 Credit point course for bachelor of medical science students
		Department of Computer Science	Course 'Computational Molecular Biology' (course code CSIS0326)	6 Credit point course for BEng degree
The Chinese University of Hong Kong (CUHK)	<a href="http://www.ust.hk">www.ust.hk</a>	School of Biomedical Science	Master of Science (M.Sc.) programme in genomics and bioinformatics	Two years programme with 24 credits
		School of Biomedical Science	Postgraduate diploma (PGD) programme in genomics and bioinformatics	One year programme 15 credit
		The Department of Computer Science and Engineering	Course 'Algorithms for Bioinformatics' (course code: CSCI3220)	For BEng of computer science students
		The Department of Computer Science and Engineering	Course 'Topics in Bioinformatics and Computational Biology' (course code: CSCI5050)	For BEng of computer science students
The Hong Kong University of Science & Technology (HKUST)	<a href="http://www.ust.hk">www.ust.hk</a>	Department of Electronic and Computer Engineering	Course 'Introduction to Bioinformatics Algorithms' (course code: ELEC 5810)	An introductory course for postgraduate programme
Hong Kong Baptist University (HKBU)	<a href="http://www.hkbu.edu.hk/">www.hkbu.edu.hk/</a>	Department of Computer Science	–	No course is available, research interest—computational biology and system biology
City University of Hong Kong (CityU)	<a href="http://www.cityu.edu.hk/">http://www.cityu.edu.hk/</a>	Department of Computer Science	Course 'Computational Biology and Bioinformatics (course code: CS4465)	3 Credit point course
The Hong Kong Polytechnic University (PolyU)	<a href="http://www.polyu.edu.hk">www.polyu.edu.hk</a>	–	–	–

amount of information produced by modern molecular biology techniques.

In September 2011, the CUHK began to offer a Master of Science (M.Sc.) programme and a postgraduate diploma (PGD) course in genomics and bioinformatics, as part-time programmes taught by the division of Genomics and Bioinformatics within the School of Biomedical Sciences. The M.Sc. programme is a two-year course carrying 24 credits while the PGD course is for 1 year with 15 credits. In both programmes, students learn fundamental subjects such as an introduction to programming, an introduction to molecular biology and genetics, biocomputing and theories and algorithms in bioinformatics. The M.Sc. programme includes some additional mandatory courses such as systems biology and genome informatics, and a research project carrying 3 credits. Two other courses related to computational biology are available in the CUHK's undergraduate programme leading to a Bachelor of

Engineering degree in computer science: 'Algorithms for Bioinformatics' (course code: CSCI3220) and 'Topics in Bioinformatics and Computational Biology' (course code: CSCI5050). Course CSCI3220 includes topics related to sequence alignment and multiple sequence alignment using different approaches such as Markov property, recursive functions, dynamic programming, FASTA and BLAST and typical clustering algorithms for microarray analysis. Course CSCI5050 includes the topics of molecular biology, data mining, data processing, sequence alignment and biological networks.

The HKUST does not have a separate programme for computational biology. However, there is one course available in the postgraduate programme of the Department of Electronic and Computer Engineering, 'Introduction to Bioinformatics Algorithms' (course code: ELEC 5810), an introductory course on computational biology at the molecular level.

The CityU, founded in 1984 as the City Polytechnic of Hong Kong, has been a fully accredited university since 1994. The Department of Computer Science offers a course entitled ‘Computational Biology and Bioinformatics’ (course code: CS4465), which aims to introduce students to the concepts and techniques used in computational biology to develop the practical skills required to solve problems in this domain.

Several workshops have been arranged in Hong Kong for training of manpower in the area of computational biology. Examples of such workshops include: (1) ‘Principles and Applications of Bioinformatics’, held on 22–23 September 1998 and organised by the Department of Biochemistry of the HKU; (2) ‘Phylogenetics and Bioinformatics’, held on 1–4 April 2008 in the School of Public Health at the CUHK; and (3) ‘Bioinformatics Algorithms and Core Technology’, held on 5 September 2012 and arranged by the Department of Computer Science and the Centre for Genomic Sciences of the HKU. In these workshops, computer science researchers were trained in how to obtain information regarding applications of the life sciences or to provide knowledge of computing for life science researchers. In Hong Kong, the objective of this type of training is to educate both young researchers and instructors. These manpower training and workshop activities were much approved and supported by the Hong Kong SAR government.

The Bioinformatics Special Interest Group (BSIG) brings people working on bioinformatics in Hong Kong together and offers an interdisciplinary environment in which to share information in this field. Group members organise regular seminars and workshops. The Asia Pacific Bioinformatics Network (APBioNET) is the oldest bioinformatics organisation in Asia. It organises the annual International Conference on Bioinformatics (InCoB), which brings together many scientists working in the field of bioinformatics. The sixth InCoB, taking place in Hong Kong on 27–30 August 2007, highlighted the growth in research excellence in the Asia Pacific region [15]. The IEEE International Conference on Bioinformatics and Biomedicine (BIBM) on 18–21 December 2010 and the IAENG International Conference on Bioinformatics (ICB’12) on 14–16 March 2012 were both held in Hong Kong. These conferences provided an international forum for scientists and researchers to exchange scientific ideas in the field.

## 5 Research in Hong Kong

Research activity in the area of computational biology is increasing day by day in Hong Kong (Fig. 3). Computational biology scientists in Hong Kong are currently inclined towards the areas of database development, sequence analysis and genome analysis. However, other prominent areas of computational biology such as structural bioinformatics,

protein networking, gene networking, drug development bioinformatics, systems biology, algorithm development and bioinformatics tools are included in the count of publications shown in Fig. 4a and give a rough indication of the number of publications originating from Hong Kong in those areas. We used the same keywords to find the number of publications in spans of 4 years from 1997 to 2012 and found a significant increase in the number of published papers originating in Hong Kong on different areas of bioinformatics (Fig. 4b).

We have searched research publications in this area in spans of 4 years, i.e. 1997–2000, 2001–2004, 2005–2008 and 2009–2012, using certain keywords. The number of publications provides a rough indication of the expansion in the field. Note, however, that some bioinformatics publications cannot be retrieved using these keywords.

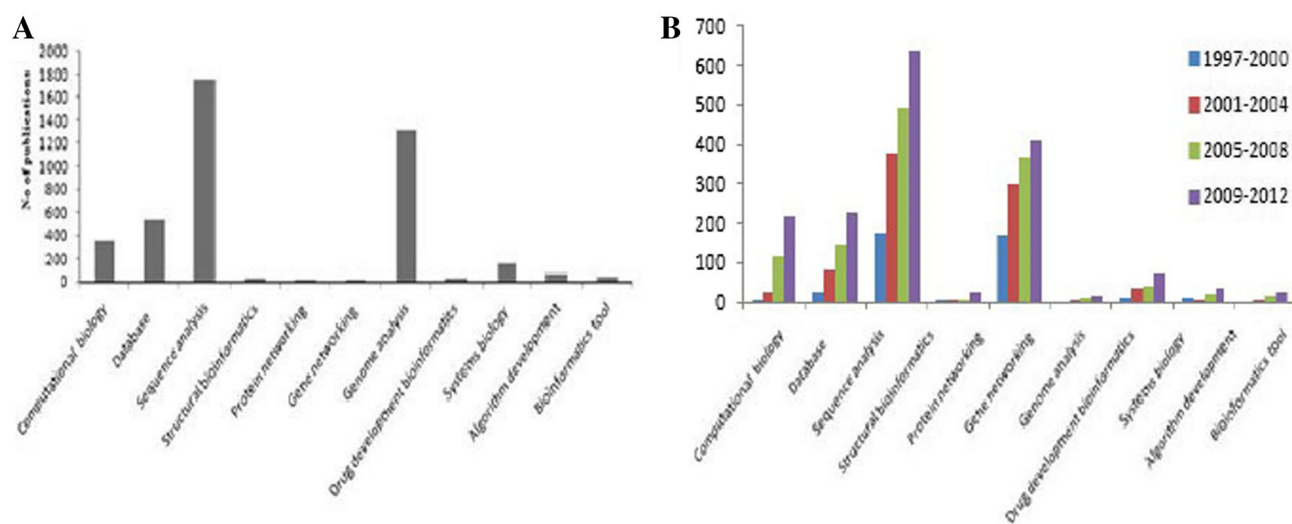
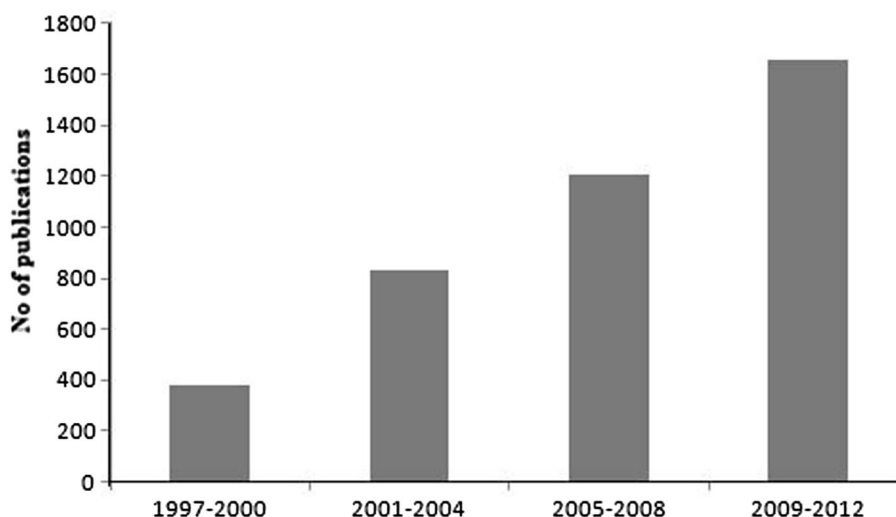
## 6 Significant Contributions of Computational Biologists

Genome analysis and sequence analysis are the main research priorities in Hong Kong, as is clear from Fig. 4. Significant contributions of Hong Kong scientists working in these areas have been noted. Professor Tsui Kwok Wing Stephen, the director of the Hong Kong Bioinformatics Centre (<http://www.hkbic.cuhk.edu.hk/>), was involved in the development of the haplotype map of the human genome as a member of the International HapMap Consortium [16]. Research laboratories in the Department of Biochemistry of the CUHK, the Department of Biochemistry and the Applied Genomics Center of the HKUST and the Genome Research Centre of the HKU were involved in the development of the second-generation human haplotype map in which researchers identified more than 3.1 million single-nucleotide polymorphisms (SNPs) [17]. Meanwhile, the Applied Genomics Center and Fok Ying Tung Graduate School of the HKUST, the Department of Biochemistry at the HKUST and the Institute of Digestive Disease at the CUHK participated in cancer genome projects as published in *Nature* [18]. Human brachydactyly type A1 (BDA1) was the first disorder recorded with autosomal-dominant type abnormalities. A study by Hong Kong researchers revealed that the BDA1 mutation (E95K) impairs the interaction of Indian hedgehog (IHH) receptor PTCH1 with the antagonist HIP1 [19]. Other examples of genome analysis and sequence analysis carried out in Hong Kong are given in Table 2.

## 7 Tools, Web Servers, Database Research

Hong Kong scientists have also been involved in the development of various bioinformatics tools. Some advanced tools include PriVar, a tool for analysing next-

**Fig. 3** Increase in ‘computational biology’ research from the Hong Kong SAR reported in PubMed references from 1997 to 2012



**Fig. 4** Number of published papers from Hong Kong in different areas of bioinformatics. **a** Total number of published papers from Hong Kong in different areas of bioinformatics. Keyword searches were performed on the PubMed database of the National Center for Biotechnology Information (NCBI). Keyword searches were carried out as ‘(Hong Kong [Affiliation]) AND (Computational biology)’; ‘(Hong Kong [Affiliation]) AND (Database)’; ‘(Hong Kong [Affiliation]) AND (Sequence analysis)’; ‘(Hong Kong [Affiliation]) AND (Structural bioinformatics)’; ‘(Hong Kong [Affiliation]) AND (Protein networking)’; ‘(Hong Kong [Affiliation]) AND (Gene networking)’; ‘(Hong Kong [Affiliation]) AND (Genome analysis)’; ‘(Hong Kong [Affiliation]) AND (Drug development bioinformatics)’; ‘(Hong Kong [Affiliation]) AND (Systems biology)’; ‘(Hong Kong [Affiliation]) AND (Algorithm development)’; ‘(Hong Kong

generation sequencing data, mutations and linkage analysis [20]; COPE, a tool for genome assembly using k-mer frequencies [21]; SOAP3, a rapid graphic processing unit (GPU)-based parallel alignment tool [22]; FetalQuant, a tool for estimating the foetal DNA concentration from maternal plasma DNA which uses maximum likelihood [23]; and GBOOST, a GPU-based tool for detecting gene-

[Affiliation]) AND (Bioinformatics tool)’. The number of publications provides a rough indication of the extent of research in the field. Note, however, that some bioinformatics publications cannot be retrieved using these keywords. **b** The increase in the number of published papers from Hong Kong in different areas of bioinformatics over time. We have grouped the period from 1997 to 2012 into spans of 4 years (1997–2000; 2001–2004; 2005–2008; 2009–2012). Keyword searches were performed on the PubMed database of the NCBI. Keyword searches were carried out in each 4-year span as, for example, ‘((Hong Kong [Affiliation]) AND Computational biology) AND (“2009/01/01” [Date-Publication]: “2012/12/12” [Date-Publication])’. The number of publications provides a rough indication of expansion in the field. Note, however, that some bioinformatics publications cannot be retrieved using these keywords

gene interactions in genome-wide case control studies [24]. Scientists from the Hong Kong Bioinformatics Centre are also working in this area. The most essential tools developed by these scientists include the ABMapper (a tool for multi-location searching and splice-junction mapping) [25], ViralFusionSeq (a tool for studying soft-clipping, read-pair and targeted de novo assembly to determine and

**Table 2** Published work from Hong Kong on genome analysis and sequence analysis by bioinformatics scientists

Organism type	Gene type/pathogen type	Type of analysis	References
Human	Locus for type 2 diabetes at 7q32 near PAX4	Study discovered rs10229583 near PAX4 as a locus for T2D in Chinese and other populations	[49]
	Mendelian disease-causing non-synonymous single nucleotide	Approximate 5 % of nsSNVs is pathogenic and carries ~22 pathogenic resulting alleles and this may lead to recessive diseases through consanguineous marriages	[50]
	Loci in or near CDKN1B, TET3, CD80, DRAM1 and ARID5B as associated with systemic lupus erythematosus in Asians	Identified genetic variations in or near CDKN1B, TET3, CD80, DRAM1 and ARID5B which are linked with the disease	[51]
	MicroRNA-218	MicroRNA-218 is related to cycle progression and apoptosis in colon cancer. This miR-218 slows down cell cycle progression and endorses apoptosis	[52]
	SNP detection	Rapid and precise SNP detection algorithm which help to analyse next-generation sequencing data	[43]
	SNP and CNVs at genome-wide scale	Study concluded 79 genes obstruct by CNVs in diseases people and recognised de novo DKK4 duplication	[53]
	CD14 gene polymorphism	Study explored the relationship between periodontitis and single polymorphic location in two genes which are DEFBI and CD14	[54]
Pathogen	Salmonella spp.	Study explored occurrence and antimicrobial resistance of Salmonella in meat harvest	[55]
	Mycobacterium tuberculosis	This experiments conclude about quick detection of <i>Mycobacteria</i> and fast exposure of drug resistance	[56]
	Rat noroviruses	Whole genome sequences	[57]
	<i>Enterobacter cloacae</i> subsp. <i>cloacae</i> strain ENHKU01	Whole genome sequences	[58]
	HIV-1 CRF07_BC variants	Detection of drug resistant mutations	[59]
	<i>Bacillus macauensis</i> ZFHKF-1	Rough sketch about genome sequence	[60]
	Influenza B virus	Structural starting point for RNA binding and homo-oligomer construction	[61]
Human coronavirus NL63	Study deals with disease variety and genetic multiplicity	[62]	

annotate human viral integration and restructure fusion) [26] and AIns (a tool for searchable and filterable sequence alignment) [27]. Further examples of Web servers or tools developed and maintained in Hong Kong are given in Table 3.

Many bioinformatics databases have been developed and maintained by Hong Kong computational biologists. Some examples are YY1TargetDB (a Web-based YY1 target loci database) [28], BSRD (a comprehensive bacterial sRNAs database) [29] and PearnBase (a searchable database for the brain coral *Platygyra carnosus*) [30]. Further examples of biological databases developed and maintained in Hong Kong are given in Table 4.

## 8 Swine Flu, Bird Flu and Respiratory Diseases

Respiratory diseases, as one of the foremost problems in Hong Kong, are a major burden [31]. Respiratory diseases and related viruses are thus important research areas in the

territory, and much landmark research has been published in well-known journals such as *Nature* and *Science*. Computational biology researchers are currently investigating the H5N1 virus, strains of which cause H5N1 avian influenza may be transferrable to humans [32]. This virus is highly pathogenic and has caused a pandemic ‘bird flu’ incident in the Hong Kong SAR. Several researchers have reported the emergence of multiple genotypes of the H5N1 virus in terrestrial poultry, leading to an outbreak of avian influenza in chickens in retail markets in Hong Kong and the molecular changes in the virus associated with this event [33]. One study published in *Nature* investigated the long-term evolution and dynamics of transmission of swine influenza A virus (SwIV) using a data set of more than 650 SwIV isolates and more than 800 swine sera from 12 years of systematic monitoring in Hong Kong, including the H1N1/2009 virus that caused a human pandemic [34]. Another study characterised a reassortant progeny of H1N1/2009 with swine viruses [35]. A study published in *Nature* on the origins and evolutionary genomics of swine-

**Table 3** Examples of Web servers or tools developed and maintained in Hong Kong

Tool Name	Function	Web address	References
GWAS3D	Tool for determination human regulatory variants through the investigation of genome-wide links, chromosome connections and histone modifications	<a href="http://jjwanglab.org/gwas3d">http://jjwanglab.org/gwas3d</a>	[63]
IPGWAS	Tool for detection of the rational thresholds in QC of GWAS datasets, relationship analysis, Manhattan plot, quantile–quantile plot and format conversion for different genetic analyses	<a href="http://sourceforge.net/projects/ipgwas/">http://sourceforge.net/projects/ipgwas/</a>	[64]
IDBA-UD	An algorithm is derived from the De Bruijn graph theory for assembling reads from single-cell sequencing or meta-genomic sequencing knowledge	<a href="http://www.cs.hku.hk/~alse/idba_ud">http://www.cs.hku.hk/~alse/idba_ud</a>	[65]
SEQanswers	An server area for collaboratively to decipher genomes and this Web server is open-access mode	<a href="http://SEQanswers.com/">http://SEQanswers.com/</a>	[66]
IGG3	A tool for complete-genome accusation and individual-level meta-analysis	<a href="http://bioinfo.hku.hk/iggweb">http://bioinfo.hku.hk/iggweb</a>	[67]
EpiRegNet	A tool for building of epigenetic dogmatic network from high-throughput gene expression	<a href="http://jjwanglab.org/EpiRegNet">http://jjwanglab.org/EpiRegNet</a>	[68]
RNASAlign	A tool for RNA structural alignment	<a href="http://www.bio8.cs.hku.hk/RNASAlign">http://www.bio8.cs.hku.hk/RNASAlign</a>	[69]
FastPval	A tool to compute extremely low P-values from empirical distribution	<a href="http://wanglab.hku.hk/pvalue">http://wanglab.hku.hk/pvalue</a>	[70]
DSHIFT	A Web server to calculate DNA chemical modifications	<a href="http://www.chem.cuhk.edu.hk/DSHIFT">http://www.chem.cuhk.edu.hk/DSHIFT</a>	[71]
mGOASVM	Multi-label analyst to calculate the sub-cellular localisation of multi-location proteins	<a href="http://bioinfo.eie.polyu.edu.hk/mGoaSvmServer/mGOASVM.html">http://bioinfo.eie.polyu.edu.hk/mGoaSvmServer/mGOASVM.html</a>	[72]

**Table 4** Examples of biological databases developed and maintained in Hong Kong

Databases name	Function of the database	Web address	References
RedoxDB	Database for protein oxidative alteration	<a href="http://biocomputer.bio.cuhk.edu.hk/RedoxDB">http://biocomputer.bio.cuhk.edu.hk/RedoxDB</a>	[73]
GWASdb	An spontaneous, multifunctional database for biologists and clinicians to investigate genetic variants and their functional conclusion	<a href="http://jjwanglab.org/gwasdb">http://jjwanglab.org/gwasdb</a>	[74]
16SpathDB	Database with 16S rRNA gene sequences of all clinically significant bacteria which are listed in the Manual of Clinical Microbiology	<a href="http://147.8.74.24/16SpathDB">http://147.8.74.24/16SpathDB</a>	[75]
OpenADAM	Web-supported data administration system for the large amount of genotype data generated from the Affymetrix GeneChip Mapping Array and Genome-Wide Human SNP Array platforms and it is open source	–	[76]
T3DB	Database that includes bacterial type III secretion process	<a href="http://biocomputer.bio.cuhk.edu.hk/T3DB/browse">http://biocomputer.bio.cuhk.edu.hk/T3DB/browse</a>	[77]
MMDDB	An incorporated medicinal resources DNA database	<a href="http://www.cuhk.edu.hk/icm/mmdbd.htm">http://www.cuhk.edu.hk/icm/mmdbd.htm</a>	[78]

origin H1N1 influenza carried out phylogenetic analysis of the gaps in genetic surveillance and applied evolutionary analysis to estimate the timescale of the origins, finding that a remixture of swine influenza lineages may have occurred years before appearing in humans [36].

## 9 Collaborative Research

Scientists in Hong Kong are actively collaborating with those in other countries such as the USA, UK, Singapore, Japan, Taiwan and India. Some important research has resulted. One study, a collaboration between researchers in

the USA, Singapore and Hong Kong, deals with a unique data set arising from surveillance of swine influenza at a Hong Kong abattoir from 1998 to 2010 and may advance understanding of the prevalence of influenza and decrease the occurrence of influenza in Hong Kong [37]. Another study aimed to understand the dissimilarity of two new *Dehalococcoides mccartyi* strains through collaboration between US and Hong Kong scientists. In this study, Lee et al. [38] performed a comparative genomics analysis via a microarray and concluded that the observed functional incongruence between the activity and core genome phylogenies of *D. mccartyi* strains is probably caused by a horizontal shift in significant reductive dehalogenase-



encoding genes. Another example of effective collaborative research is the inclusion of different species in the Molecular Ecology Resources Database in which several countries, such as Australia, Austria, the USA, Brazil, Germany, Sweden, Taiwan, China, Canada, India and Hong Kong, participated [39, 40].

The authors of this paper are themselves performing collaborative research. Three of the authors (George, Chakraborty and Zhu) have jointly published research findings on the effects of deleterious non-synonymous SNPs in the binding adaptability of flavopiridol with cyclin-dependent kinase 7 (CDK7), a one-cell-cycle regulatory protein [41]. Another example of collaboration between India and Hong Kong is this review paper concerning computational biology in the Hong Kong SAR; bioinformatics scientists from both countries have critically analysed the activity in and the status of computational biology in the territory.

## 10 Some Eminent Scientists

Eminent Hong Kong researchers in the field of computational biology include Professor Tsui Kwok Wing Stephen, one of the authors of this review and the director of the Hong Kong Bioinformatics Centre who was a member of the International HapMap Consortium developing the haplotype map of the human genome; two papers on that work were published in *Nature* in 2005 and 2007 [16, 17]. Professor Tsui's group worked on genomic-sequence variations and the epidemiology of severe acute respiratory syndrome (SARS) [42]. Dr. Jun Wang from the Department of Biochemistry at the HKU is working on computational and transcriptional genomics, structural bioinformatics, SNPs and copy number variation (CNV). He developed an SNP detection algorithm for next-generation sequencing data [43] and also worked on a genome-wide association study on alleles in the FGFR2 gene that is associated with risk of breast cancer [44]. Ascertaining the dynamic nature of gene regulation is a significant challenge in systems biology, an area of research interest in Hong Kong. The corresponding author of this review is working on reconstructing dynamic gene regulatory networks for human cancer, to unravel the dynamic mechanism of cancer development [45]. Dr. Yip Yuk Lap Kevin from the School of Life Sciences at CUHK is researching the use of computational methods to study biological and medical phenomena and networks, and has published work in the journals *Nature* and *Bioinformatics* [46, 47]. Other eminent Hong Kong researchers have been working in the field of computational biology for decades; an attempt to outline the activities of these researchers, to provide a bird's-eye view of their research activities, is given in Table 5. We have also attempted to determine the

proportion of total research in Hong Kong that is related to computational biology, finding that publications in the area of bioinformatics comprise about 9 % of total publications (Fig. 5).

The Hong Kong SAR government provides several sources of research funding, such as the General Research Fund (GRF), Collaborative Research Fund (CRF) and Theme-based Research Scheme administered through the Research Grants Council (RGC) (<http://cerg1.ugc.edu.hk/>); the ITF administered through the ITC (<http://www.itc.gov.hk/>); and the Health and Medical Research Fund (HMRF) (formerly known as the Health and Health Services Research Fund—HHSRF—and the Research Fund for the Control of Infectious Diseases—RFCID) administered through the Research Fund Secretariat of the Food and Health Bureau.

The percentage of publications provides a rough indication of the amount of research being carried out. Here, to search for total publications, we used the search term '(Hong Kong [affiliation])'. For bioinformatics publications, we used the total number of publications from Fig. 3.

## 11 Activities of Bioinformatics Companies

Hong Kong relies on medical devices being imported to satisfy the territory's rising demand for advanced health care. A major focus of the Hong Kong SAR government, therefore, is assisting companies to fulfil the need for medical products. Recent trends in Hong Kong's population have shown a shift of disease types towards malignant neoplasms, heart disease, pneumonia, cerebrovascular disease, etc. There is a resulting increase in the demand for high-tech products and services that use biotechnology and bioinformatics tools. In response to the outbreak of SARS, many biotechnological companies were established to develop diagnostic kits. One such project involves the development of a biochip by Dr. Yu Cheung-Hoi Albert of the HKUST, which was highlighted in *Science* magazine [48]. This 'lab-on-a-chip' rapidly diagnoses emerging pathogens, including influenza viruses; Yu subsequently established the company Hai Kang Life to manufacture these DNA chips. Cluster Technology Limited (ClusterTech) is a computing technology company in Hong Kong which provides solutions for bioinformatics research that combine high-performance computing and cloud computing technology; several other companies have also been established in this area. To assist with world-class infrastructure for such companies, the Hong Kong Science and Technology Parks Corporation (HKSTPC) was established by the Hong Kong SAR government in May 2001 to provide core competency in manufacturing technology, biotechnology, information technology (including computational biology), environmental technology, management systems, etc. The

**Table 5** Some well-known researchers, their affiliations and areas of research interest

Prominent Scientists/researcher Name	Affiliation	Area of research interest	References
Professor Stephen Tsui	School of Biomedical Sciences, The Chinese University of Hong Kong; and Director of the Hong Kong Bioinformatics Centre, Hong Kong SAR	Human cancers and gene regulation, changes in microbiota in human diseases, bioinformatics	[16, 17, 42]
Prof. Cheah Kathy	Department of Biochemistry, HKU	Working on the directive and function of genes, disease machinery especially skeletal system, human degenerative skeletal disorders and the inner ear	[79, 80]
Dr. Yang Wanling	Department of Biochemistry, HKU	Working on genomics data assessment and statistical genetics using bioinformatics especially in Asian populations	[81, 82]
Prof. Francis Chin	Department of Biochemistry, HKU	Working on design and analysis of algorithms in the bioinformatics area	[65, 83]
Dr. Siu-Ming Yiu	Department of Biochemistry HKU	Working on bioinformatics-related algorithms	[84]
Dr. Hing-Fung Ting	Department of Biochemistry HKU	Working on bioinformatics-related algorithms	[85, 86]
Dr. Chan Ting Fung	School of Life Sciences of CUHK	Pursuing research on bioinformatics analysis of microbial genomes and transcriptomes, as well as genetic and genomic analysis of complex human diseases	[87, 88]
Dr. Yip Yuk Lap Kevin	School of Life Sciences of CUHK	Working on using computational methods to study biological and medical phenomena	[46, 47]
Prof. Yang Qiang	Department of Computer Science and Engineering, HKUST	Algorithms and software tools on computational biology	[89, 90]
Prof. Tang Lei-Han	Department of Physics, HKUST	Working on unravelling the underlying mechanism of conformational transformations of biopolymers as well as modelling the metabolic network of a biological system	[91, 92]
Prof. Ng Michael	Department of Mathematics, HKBU	Theoretical study on bioinformatics algorithm	[93, 94]
Dr. Zhu Hailong	Department of Computer Science, HKBU	Hypothetical and computational methods of rebuild gene regulatory networks as well as transcriptional regulatory networks	[45, 95, 96]
Prof. Wang Lusheng	Department of Computer Science in CityU	Working on the algorithms and networks of computational molecular biology	[97, 98]

HKSTPC manages three industrial estates and mentors technology-based companies through its incubation programme which assists with business development for potential new entrepreneurs.

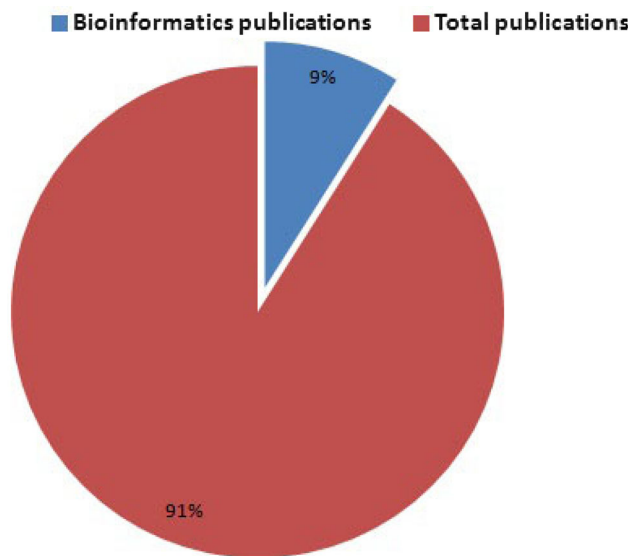
## 12 Government Role

The Hong Kong SAR government plays a very supportive role in promoting innovation and the development of technology. The government is extremely keen to create an environment which promotes such innovation and development and demonstrates a particular interest in biotechnology, thus fostering favourable conditions for computational biologists. The government supports the innovation and technology support, university–industry collaboration, general support and small entrepreneur research assistance programmes. At the beginning of 2012,

for example, 2746 projects were supported from a fund of HK\$6.4 billion, with many of the funded projects being related to biotechnology (around 10 %) and information technology (around 20 %), jointly comprising 30 % of the total funding in the territory. The Hong Kong SAR government generally supports fundamental research, applied R&D, technology transfer and technological entrepreneurship, through which it fosters a culture of innovation and a technological environment in the territory.

## 13 Future Challenges and Concluding Remarks

In the past few years, Hong Kong has seen tremendous growth in bioinformatics research, which has been extended to different fields of biology. However, compared with other developed countries around the world, Hong Kong still has some way to go in this field. Research in the life



**Fig. 5** Comparison between bioinformatics publications and total publications originating from Hong Kong, retrieved from the PubMed database of the NCBI

sciences with the aid of computational biology is occurring on a large scale in Hong Kong and is still gaining impetus. Collaborative research activities between bioinformaticians and laboratory biologists at both a national and an international level are also increasing rapidly. Such collaborative research groups analyse biological data ranging from genes to proteins, attempt to solve important biological questions together with laboratory scientists, make various predictions and validate different hypotheses. Despite this positive outlook, however, several challenges exist. More national centres for bioinformation, similar to the National Center for Biotechnology Information (NCBI) in the USA, are needed to provide a central depository and information source for biological data and tools.

During the past few years, development of computational biology has been initiated in Hong Kong through various activities. However, many challenges lie ahead for computational biologists and the Hong Kong SAR government if the territory is to be established as a globally important bioinformatics research centre. The government needs to act to encourage more research projects, teaching modules and conferences and the provision of research grants and support to attract larger numbers to the field of computational biology. Research must be sustained as a key area of investment to improve efficiency and competitiveness in meeting Hong Kong's needs, such as in the medical sector. To identify and support the existing researchers in the field of computational biology, more professional societies and journals indexed in the Science Citation Index should be initiated to drive Hong Kong computational biology towards new heights. This will

benefit both national scientists and those in neighbouring Asian countries. Finally, venture capital would be extremely beneficial in the development of bioinformatics companies to assist Hong Kong to become a world-famous centre of computational biology.

**Acknowledgments** This work was supported by the Research Grants Council of Hong Kong (212111 and 212631), Faculty Research Grant (12-13/061) and partially supported by the National Natural Science Foundation of China (61134013 and 91029301). The authors take this opportunity to thank the management of Galgotias University and VIT University for providing the facilities and encouragement to carry out this work.

#### Compliance with Ethical Standards

**Conflict of interest** The authors declare no conflicts of interest.

#### References

1. Yu TFL (1997) *Entrepreneurship and the economic development of Hong Kong*. Routledge, London, p 240
2. Russell A (2006) *The top 10 of everything 2007*. Hamlyn, London, p 78
3. Gascoigne B, Gascoigne C (2003) *The dynasties of China: a history*. Running Press, Philadelphia
4. Porter B (2004) *The lion's share: a short history of British imperialism, 1850–2004*. Addison-Wesley Longman, Boston
5. Cohen WI, Li Z (1997) *Hong Kong under Chinese rule: the economic and political implications of reversion*. Cambridge University Press, Cambridge
6. Mark CK (2004) *Hong Kong and the cold war: Anglo-American relations 1949–1957*. Oxford University Press, Oxford
7. Kong L (2012) *Improbable art: the creative economy and sustainable cluster development in a Hong Kong industrial district*. *Eurasian Geogr Econ* 53:182–196
8. Enright MJ, Scott EE, Dodwell D (2011) *The Hong Kong advantage*. Oxford University Press, Oxford
9. Young A (1992) *A tale of two cities: factor accumulation and technical change in Hong Kong and Singapore*. In: *NBER macroeconomics annual 1992*. MIT Press, vol 7, pp 13–64
10. Phang HS (2011) *Population aging and productivity in Asian countries*. Japan, Asian Productivity Organization, p 142
11. Suen WM, Chick KW (1988) *Use of microcomputer for histopathology: system using IBM PC and dBaseIII*. *J Clin Pathol* 41(2):220–222
12. Ling JM, Hui YW, French GL (1988) *Evaluation of the Microbact-24E bacterial identification system*. *J Clin Pathol* 41(8):910–914
13. Smith DK, Xue H (1998) *A major component approach to presenting consensus sequences*. *Bioinformatics* 14(2):151–156
14. Chau FT, Chan TP, Wang J (1998) *TLCQA: quantitative study of thin-layer chromatography*. *Bioinformatics* 14(6):540–541
15. Ranganathan S, Gribskov M, Tan TW (2008) *Bioinformatics research in the Asia Pacific: a 2007 update*. *BMC Bioinformatics* 9:S1
16. *The International HapMap Consortium* (2005) *A haplotype map of the human genome*. *Nature* 437:1299–1320
17. *The International HapMap Consortium* (2007) *A second generation human haplotype map of over 3.1 million SNPs*. *Nature* 449:851–861
18. *The International Cancer Genome Consortium* (2010) *International network of cancer genome projects*. *Nature* 464:993–998

19. Gao B, Hu J, Stricker S et al (2009) A mutation in *Ihh* that causes digit abnormalities alters its signalling capacity and range. *Nature* 458(7242):1196–1200
20. Zhang L, Zhang J, Yang J et al (2013) PriVar: a toolkit for prioritizing SNVs and indels from next-generation sequencing data. *Bioinformatics* 29(1):124–125
21. Liu B, Yuan J, Yiu SM et al (2012) COPE: an accurate k-mer-based pair-end reads connection tool to facilitate genome assembly. *Bioinformatics* 28(22):2870–2874
22. Liu CM, Wong T, Wu E et al (2012) SOAP3: ultra-fast GPU-based parallel alignment tool for short reads. *Bioinformatics* 28(6):878–879
23. Jiang P, Chan KC, Liao GJ et al (2012) FetalQuant: deducing fractional fetal DNA concentration from massively parallel sequencing of DNA in maternal plasma. *Bioinformatics* 28(22):2883–2890
24. Yung LS, Yang C, Wan X et al (2011) GBOOST: a GPU-based tool for detecting gene–gene interactions in genome-wide case control studies. *Bioinformatics* 27(9):1309–1310
25. Lou SK, Ni B, Lo LY et al (2011) ABMapper: a suffix array-based tool for multi-location searching and splice-junction mapping. *Bioinformatics* 27:421–422
26. Li JW, Wan R, Yu CS et al (2013) ViralFusionSeq: accurately discover viral integration events and reconstruct fusion transcripts at single-base resolution. *Bioinformatics* 29(5):649–651
27. Leung RK, Tsui SK (2013) Alns: a new searchable and filterable sequence alignment format. *Int J Data Min Bioinform* 7(2):135–145
28. Guo AM, Sun K, Su X et al (2013) TargetDB: an integral information resource for Yin Yang 1 target loci. *Database (Oxford)* 2013:bat007
29. Li L, Huang D, Cheung MK et al (2013) BSRD: a repository for bacterial small regulatory RNA. *Nucleic Acids Res* 41:D233–D238
30. Sun J, Chen Q, Lun JC et al (2013) PearnBase: development of a transcriptomic database for the brain coral *Platygyra carnosus*. *Mar Biotechnol (NY)* 15(2):244–251
31. Chan-Yeung M, Lai CK, Chan KS, Hong Kong Thoracic Society et al (2008) The burden of lung disease in Hong Kong: a report from the Hong Kong Thoracic Society. *Respirology* 13(4):S133–S165
32. Yen HL, Peiris JS (2012) Virology: bird flu in mammals. *Nature* 486(7403):332–333
33. Guan Y, Peiris JS, Lipatov AS et al (2002) Emergence of multiple genotypes of H5N1 avian influenza viruses in Hong Kong SAR. *Proc Natl Acad Sci USA* 99:8950–8955
34. Vijaykrishna D, Smith GJ, Pybus OG et al (2011) Long-term evolution and transmission dynamics of swine influenza A virus. *Nature* 473(7348):519–522
35. Vijaykrishna D, Poon LL, Zhu HC et al (2010) Reassortment of pandemic H1N1/2009 influenza A virus in swine. *Science* 328:1529
36. Smith GJ, Vijaykrishna D, Bahl J et al (2009) Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 459(7250):1122–1125
37. Streliaff CC, Vijaykrishna D, Riley S et al (2013) Inferring patterns of influenza transmission in swine from multiple streams of surveillance data. *Proc Biol Sci* 280(1762):20130872
38. Lee PK, Cheng D, West KA et al (2013) Isolation of two new *Dehalococcoides mccartyi* strains with dissimilar dechlorination functions and their characterization by comparative genomics via microarray analysis. *Environ Microbiol* 15(8):2293–2305
39. Molecular Ecology Resources Primer Development Consortium, Andree K, Axtner J, Bagley MJ et al (2010) Permanent genetic resources added to Molecular Ecology Resources Database 1 April 2010–31 May 2010. *Mol Ecol Resour* 10(6):1098–1105
40. Molecular Ecology Resources Primer Development Consortium, Almany GR, De Arruda MP, Arthofer W et al (2009) Permanent genetic resources added to Molecular Ecology Resources Database 1 May 2009–31 July 2009. *Mol Ecol Resour* 9(6):1460–1466
41. George Priya Doss C, Nagasundaram N, Chakraborty C et al (2013) Extrapolating the effect of deleterious nsSNPs in the binding adaptability of flavopiridol with CDK7 protein: a molecular dynamics approach. *Hum Genomics* 7(1):10
42. Tsui SK, Chim SS, Lo YM, Chinese University of Hong Kong Molecular SARS Research Group (2003) Coronavirus genomic-sequence variations and the epidemiology of the severe acute respiratory syndrome. *N Engl J Med* 349(2):187–188
43. Xu F, Wang W, Wang P et al (2012) A fast and accurate SNP detection algorithm for next-generation sequencing data. *Nat Commun* 3:1258
44. Hunter DJ, Kraft P, Jacobs KB et al (2007) A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* 39:870–874
45. Zhu HL, Rao RSP, Zeng T et al (2012) Reconstructing dynamic gene regulatory networks from sample-based transcriptional data. *Nucleic Acids Res* 40(21):10657–10667
46. Gerstein MB, Kundaje A, Hariharan M et al (2012) Architecture of the human regulatory network derived from ENCODE data. *Nature* 489(7414):91–100
47. Yip KY, Gerstein M (2009) Training set expansion: an approach to improving the reconstruction of biological networks from limited and uneven reliable interactions. *Bioinformatics* 25(2):243–250
48. Stone R (2011) Biotechnology. Lab-on-a-chip maker looks to put Hong Kong on biotech map. *Science* 332:1497
49. Ma RC, Hu C, Tam CH et al (2013) Genome-wide association study in a Chinese population identifies a susceptibility locus for type 2 diabetes at 7q32 near *PAX4*. *Diabetologia* 56(6):1291–1305
50. Li MX, Kwan JS, Bao SY et al (2013) Predicting mendelian disease-causing non-synonymous single nucleotide variants in exome sequencing studies. *PLoS Genet* 9(1):e1003143
51. Yang W, Tang H, Zhang Y et al (2013) Meta-analysis followed by replication identifies loci in or near *CDKN1B*, *TET3*, *CD80*, *DRAM1*, and *ARID5B* as associated with systemic lupus erythematosus in Asians. *Am J Hum Genet* 92(1):41–51
52. He X, Dong Y, Wu CW et al (2013) MicroRNA-218 inhibits cell cycle progression and promotes apoptosis in colon cancer by downregulating *BMI1* polycomb ring finger oncogene. *Mol Med* 18:1491–1498
53. Wong EH, Cui L, Ng CL et al (2013) Genome-wide copy number variation study in anorectal malformations. *Hum Mol Genet* 22(3):621–631
54. Loo WT, Bai LJ, Fan CB et al (2012) Clinical application of human  $\beta$ -defensin and *CD14* gene polymorphism in evaluating the status of chronic inflammation. *J Transl Med* 10(1):S9
55. Wong MH, Chen S (2013) First detection of *oqxAB* in *Salmonella* spp. isolated from food. *Antimicrob Agents Chemother* 57(1):658–660
56. Yam WC, Siu KH (2013) Rapid identification of mycobacteria and rapid detection of drug resistance in *Mycobacterium tuberculosis* in cultured isolates and in respiratory specimens. *Methods Mol Biol* 943:171–199
57. Tse H, Chan WM, Lam CS et al (2012) Complete genome sequences of novel rat noroviruses in Hong Kong. *J Virol* 86(22):12435–12436
58. Liu WY, Chung KM, Wong CF et al (2012) Complete genome sequence of the endophytic *Enterobacter cloacae* subsp. *cloacae* strain ENHKU01. *J Bacteriol* 194(21):5965
59. Wu H, Zhang HJ, Zhang XM (2012) Identification of drug resistant mutations in HIV-1 CRF07\_BC variants selected by nevirapine in vitro. *PLoS One* 7(9):e44333

60. Cai L, Zhang T (2012) Genome of *Bacillus macauensis* ZFHKF-1, a long-chain-forming bacterium. *J Bacteriol* 194(17):4780
61. Ng AK, Lam MK, Zhang H et al (2012) Structural basis for RNA binding and homo-oligomer formation by influenza B virus nucleoprotein. *J Virol* 86(12):6758–6767
62. Leung TF, Chan PK, Wong WK et al (2012) Human coronavirus NL63 in children: epidemiology, disease spectrum, and genetic diversity. *Hong Kong Med J* 18(2):27–30
63. Li MJ, Wang LY, Xia Z et al (2013) GWAS3D: detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications. *Nucleic Acids Res*. doi:10.1093/nar/gkt456
64. Fan YH, Song YQ (2012) IPGWAS: an integrated pipeline for rational quality control and association analysis of genome-wide genetic studies. *Biochem Biophys Res Commun* 422(3):363–368
65. Peng Y, Leung HC, Yiu SM et al (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28(11):1420–1428
66. Li JW, Schmieder R, Ward RM et al (2012) SEQanswers: an open access community for collaboratively decoding genomes. *Bioinformatics* 28(9):1272–1273
67. Li MX, Jiang L, Kao PY et al (2009) IGG3: a tool to rapidly integrate large genotype datasets for whole-genome imputation and individual-level meta-analysis. *Bioinformatics* 25(11):1449–1450
68. Wang LY, Wang P, Li MJ et al (2011) EpiRegNet: constructing epigenetic regulatory network from high throughput gene expression data for humans. *Epigenetics* 6(12):1505–1512
69. Wong TK, Wan KL, Hsu BY et al (2011) RNASAlign: RNA structural alignment system. *Bioinformatics* 27(15):2151–2152
70. Li MJ, Sham PC, Wang J (2010) FastPval: a fast and memory efficient program to calculate very low P-values from empirical distribution. *Bioinformatics* 26(22):2897–2899
71. Lam SL (2007) DSHIFT: a web server for predicting DNA chemical shifts. *Nucleic Acids Res* 35:W713–W717
72. Wan S, Mak MW, Kung SY (2012) mGOASVM: Multi-label protein subcellular localization based on gene ontology and support vector machines. *BMC Bioinformatics* 13:290
73. Sun MA, Wang Y, Cheng H, Zhang Q, Ge W, Guo D (2012) RedoxDB: a curated database for experimentally verified protein oxidative modification. *Bioinformatics* 28(19):2551–2552
74. Li MJ, Wang P, Liu X et al (2012) GWASdb: a database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res* 40:D1047–D1054
75. Woo PC, Teng JL, Yeung JM et al (2011) Automated identification of medically important bacteria by 16S rRNA gene sequencing using a novel comprehensive database, 16SpathDB. *J Clin Microbiol* 49(5):1799–1809
76. Yeung JM, Sham PC, Chan AS et al (2008) OpenADAM: an open source genome-wide association data management system for Affymetrix SNP arrays. *BMC Genomics* 9:636
77. Wang Y, Huang H, Sun M et al (2012) T3DB: an integrated database for bacterial type III secretion system. *BMC Bioinformatics* 13:66
78. Lou SK, Wong KL, Li M et al (2010) An integrated web medicinal materials DNA database: MMDBD (Medicinal Materials DNA Barcode Database). *BMC Genomics* 11:402
79. Scott CE, Wynn SL, Sesay A et al (2010) SOX9 induces and maintains neural stem cells. *Nat Neurosci* 13(10):1181–1189
80. Leung VY, Gao B, Leung KK et al (2011) SOX9 governs differentiation stage-specific gene expression in growth plate chondrocytes via direct concomitant transactivation and repression. *PLoS Genet* 7(11):e1002356
81. Yang W, Shen N, Ye D-Q et al (2010) Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. *PLoS Genet* 6(2):e1000841
82. Yang W, Zhao M, Hiranarn N et al (2009) ITGAM is associated with disease susceptibility and renal nephritis of systemic lupus erythematosus in Hong Kong Chinese and Thai. *Hum Mol Genet* 18:2063–2070
83. Wang Y, Henry CM, Leung SM et al (2012) MetaCluster 5.0: a two-round binning approach for metagenomic data for low-abundance species in a noisy sample. *Bioinformatics* 28(18):i356–i362
84. Wong TK, Yiu SM (2012) Structural alignment of RNA with triple helix structure. *J Comput Biol* 19(4):365–378
85. Kao MY, Lam TW, Przytycka TM et al (1997) General techniques for comparing unrooted evolutionary trees. In: *Proceedings of the twenty-ninth annual ACM symposium on theory of computing*, pp 54–65
86. Kao MY, Lam TW, Sung WK, et al (1999) A decomposition theorem for maximum weight bipartite matchings with applications to evolutionary trees. In: *Algorithms-ESA'99*. Berlin, Springer, pp 438–449
87. Li JW, Bolser D, Manske M et al (2013) The NGS WikiBook: a dynamic collaborative online training effort with long-term sustainability. *Brief Bioinform* 14(5):548–555
88. Yu CS, Yim KY, Tsui SK et al (2012) Complete genome sequence of *Bacillus subtilis* strain QB928, a strain widely used in *B. subtilis* genetic studies. *J Bacteriol* 194(22):6308–6309
89. Zhou Y, Tang M, Pan W et al (2013) Bird flu outbreak prediction via satellite tracking. *IEEE Intell Syst* 29(4):10–17
90. Pan W, Yang Q (2013) Transfer learning in heterogeneous collaborative filtering domains. *Artif Intell J* 197:39–55
91. Xiong LP, Ma YQ, Tang LH (2009) Attenuation of transcriptional bursting in mRNA transport. *Phys Biol* 7(1):016005
92. Fu X, Tang LH, Liu C et al (2012) Stripe formation in bacterial systems with density-suppressed motility. *Phys Rev Lett* 108(19):198102
93. Liu Y, Ng M (2010) Shrunk methodology to genome-wide SNPs selection and construction of SNPs networks. *BMC Syst Biol* 4(2):S5
94. Liu Y, Lee YF, Ng MK (2011) SNP and gene networks construction and analysis from classification of copy number variations data. *BMC Bioinformatics* 12(Suppl 5):S4
95. Chow K, Zhu HL, Kuo WP (2010) Error margin analysis for feature gene extraction. *BMC Bioinformatics* 11:241
96. George Priya Doss C, Chakraborty C, Haneef SAS, NagaSundaram N, Chen L, Zhu H (2014) Evolution- and structure-based computational strategy reveals the impact of deleterious missense mutations on MODY 2 (maturity-onset diabetes of the young, type 2). *Theranostics* 4(4):366–385
97. Guo F, Wang L (2012) Computing the protein binding sites. *BMC Bioinformatics* 13(Suppl 100):S2
98. Deng F, Cui W, Wang L (2013) A highly accurate heuristic algorithm for the haplotype assembly problem. *BMC Genomics* 14(Suppl 2):S2