

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



تجزیه و تحلیل داده‌های RNA-seq

همراه با لوح فشرده آموزش گام به گام

www.ketab.ir

دکتر علیرضا سیفی

عضو هیئت علمی دانشگاه فردوسی مشهد

مهندس محمدرضا رضائی

سرشناسه:

عنوان و نام پدیدآور:

سیفی، علیرضا، ۱۳۵۶-

تجزیه و تحلیل داده‌های RNA-seq همراه با لوح فشرده آموزش گام‌به‌گام / علیرضا سیفی، محمدرضا رضایی؛ ویراستار علمی امین میرشمسی کاخکی؛ ویراستار ادبی هانیه اسدیپور فعال مشهد.

مشخصات نشر:

مشهد: دانشگاه فردوسی مشهد، ۱۴۰۰.

مشخصات ظاهری:

۱۲۸ ص.

فروست:

انتشارات دانشگاه فردوسی مشهد؛ ۷۹۰.

شابک:

ISBN: 978-964-386-483-5

وضعیت فهرست‌نویسی:

فیا.

موضوع:

آ. ان. ا.

موضوع:

نوکلئوتیدها -- توالی

شناسه افزوده:

رضایی، محمدرضا، ۱۳۷۳-

شناسه افزوده:

میرشمسی کاخکی، امین، ۱۳۶۵-، ویراستار

شناسه افزوده:

دانشگاه فردوسی مشهد، انتشارات.

رده‌بندی کنگره:

QP۶۲۳

رده‌بندی دیویی:

۵۷۲/۸۸

شماره کتابشناسی ملی:

۷۶۱۴۲۶۹

RNA

Nucleotide sequence

تجزیه و تحلیل داده‌های RNA-seq

همراه با لوح فشرده آموزش گام‌به‌گام

پدیدآورندگان:

دکتر علیرضا سیفی؛ مهندس محمدرضا رضایی

ویراستار علمی:

دکتر امین میرشمسی کاخکی

ویراستار ادبی:

هانیه اسدیپور فعال مشهد

مشخصات:

وزیری، ۲۰۰ نسخه، چاپ اول، پاییز ۱۴۰۰

چاپ و صحافی:

چاپخانه دقت

بها:

۳۲۰،۰۰۰ ریال

حق چاپ برای انتشارات دانشگاه فردوسی مشهد محفوظ است.

مواکز پخش:

فروشگاه و نمایشگاه کتاب پردیس: مشهد، میدان آزادی، دانشگاه فردوسی مشهد، جنب سلف یاس

تلفن: ۳۸۸۳۳۷۲۷-۳۸۸۰۲۶۶۶ (۰۵۱)

مؤسسه کتابیران: تهران، خیابان کارگر جنوبی، خیابان لبافی‌نژاد، بین خیابان فروردین و اردیبهشت،

شماره ۲۳۸، تلفن: ۶۶۴۸۴۷۱۵-۶۶۴۹۴۴۰۹ (۰۲۱)

مؤسسه دانشیران: تهران، خیابان انقلاب، خیابان منیری جاوید (اردیبهشت) نیش خیابان نظری، شماره ۱۴۲

تلفکس: ۶۶۴۰۰۲۲۰-۶۶۴۰۰۱۴۴ (۰۲۱)

<http://press.um.ac.ir>

Email: press@um.ac.ir



انتشارات

۷۹۰

فهرست مطالب

پیشگفتار	۷
فصل ۱. تاریخچه پیشرفت‌ها در ژنتیک، بیولوژی مولکولی و بیوانفورماتیک	۹
نسل دوم روش‌های توالی‌یابی	۱۲
نسل سوم روش‌های توالی‌یابی	۱۵
ابزارهای آنالیز داده‌های NGS	۱۶
منابع	۱۹
فصل ۲. آشنایی با اصول کامپیوتر برای پژوهشگران علوم زیستی	۲۱
انتخاب سیستم کامپیوتری مناسب برای آنالیز داده‌های NGS	۲۲
حافظه‌های ذخیره‌سازی ثانویه	۲۴
ریزپردازنده مرکزی یا CPU	۲۵
حافظه اصلی یا RAM	۲۶
منابع	۲۶
فصل ۳. سیستم عامل لینوکس	۲۷
تاریخچه نرم‌افزارهای با دسترسی آزاد و پیدایش لینوکس	۲۷
نصب سیستم عامل Ubuntu	۲۸
دستورات در Ubuntu	۲۹

۳۳	کلیدهای ترکیبی در خط فرمان لینوکس
۳۴	منابع
۳۵	فصل ۴. محیط نرم‌افزاری و زبان برنامه‌نویسی R
۳۷	نصب R
۳۹	توابع R
۴۰	ساده‌سازی با ایجاد متغیر
۴۰	پوشه کاری و تغییر آن
۴۱	داده‌ها در R
۴۲	ساختار داده‌ها در R
۴۵	روش‌های وارد کردن داده‌ها به R
۴۷	عملیات ریاضی روی وکتورها
۴۷	مقایسات منطقی
۴۸	عملگرهای منطقی
۴۹	ترسیم نمودارهای آماری در R
۵۰	نمودار هیستوگرام
۵۱	بسته نرم‌افزاری ggplot2
۵۵	برخی دیگر از توابع مهم R
۵۶	نوشتن برنامه در R
۵۸	منابع
۵۹	فصل ۵. RNA-seq: انقلابی در مطالعه ترنسکرپتوم
۵۹	روش‌های ارزیابی بیان ژن
۶۱	طراحی آزمایشات RNA-seq
۶۵	روش‌های نوین RNA-seq
۶۶	استفاده از داده‌های RNA-seq موجود در NCBI

پیشگفتار

فناوری‌های نوین توالی‌یابی DNA که با عنوان متداول NGS شناخته می‌شوند، تحول شگرفی در رشته‌های مختلف علوم زیستی ایجاد کرده و امکان مطالعه سریع‌تر و جامع‌تر اساس ژنتیکی و مولکولی پدیده‌های زیستی را فراهم آورده‌اند. هم‌راستا با موفقیت‌های چشمگیر سخت‌افزاری در فناوری‌های توالی‌یابی DNA، پیشرفت‌های اساسی نیز در تولید ابزارهای تجزیه و تحلیل داده‌های حاصل از این فناوری‌ها حاصل شده است.

چالش اصلی در تجزیه و تحلیل داده‌های NGS حجم بالا و پیچیدگی این داده‌هاست که بهره‌برداری مناسب از آن‌ها نیازمند استفاده از ابزارهای بیوانفورماتیکی کارا و دقیق است. این ابزارهای بیوانفورماتیکی معمولاً توسط متخصصان علوم ریاضی، کامپیوتر و بیوانفورماتیک ساخته می‌شوند و در غالب موارد بهره‌برداری از آن‌ها برای متخصصان علوم زیستی که اطلاعات کافی از علوم کامپیوتری ندارند، چالش‌برانگیز است. معمولاً در مراکز تحقیقاتی بویا در دنیا یک یا چند متخصص بیوانفورماتیک عهده‌دار تجزیه و تحلیل داده‌های NGS تولیدشده توسط محققان زیست‌شناسی هستند و بنابراین تمام محققان نیازمند فراگیری این روش‌های تجزیه و تحلیل نیستند. لیکن در ایران به دلیل اینکه هنوز توسعه یافتگی مطلوبی در مراکز تحقیقاتی رخ نداده است و کماکان تشکیل تیم‌های تحقیقاتی با تخصص‌های گوناگون میسر نیست، محققان علوم زیستی نیاز دارند که اشراف نسبی بر روش‌های تجزیه و تحلیل NGS داشته باشند.

مطالعه ترنسکریپتوم (کل محتوای RNA سلول) با استفاده از روش‌های توالی‌یابی RNA (که با عنوان RNA-seq شناخته می‌شود) یکی از قدرتمندترین روش‌های مطالعه پدیده‌های زیستی برای شناسایی مکانیسم‌های مولکولی و ژنتیکی کنترل‌کننده این پدیده‌هاست. ابزارهای بیوانفورماتیکی بسیار متنوعی برای تجزیه و تحلیل داده‌های RNA-seq ارائه شده است و ابزارهای جدید نیز با سرعت در حال توسعه و ارائه است. هدف از گردآوری کتاب حاضر فراهم آوردن مجموعه‌ای است برای محققان علوم زیستی که الزاماً اطلاعات گسترده‌ای از بیوانفورماتیک ندارند، ولی نیازمند تجزیه و تحلیل داده‌های RNA-seq هستند.

در این کتاب مباحث نظری مورد نیاز در مورد RNA-seq و اصول پایه و مقدماتی کامپیوتر توضیح داده می‌شود. سپس روش‌های استاندارد تجزیه و تحلیل داده‌های RNA-seq که در حال حاضر بیشترین کاربرد را در بین پژوهشگران دارند، معرفی می‌شوند. با استفاده از داده‌های واقعی کل مسیرهای تجزیه و تحلیل به صورت گام به گام توضیح داده می‌شوند، به نحوی که مخاطب با در اختیار داشتن این کتاب بتواند کل مسیر تجزیه و تحلیل RNA-seq را روی کامپیوتر شخصی خود تمرین کند. کلیه نرم‌افزارها، داده‌های مورد استفاده

و خروجی‌های موردانتظار از اجرای هر کدام از مراحل تجزیه و تحلیل در قالب یک لوح فشرده همراه کتاب در اختیار خوانندگان قرار می‌گیرد. تلاش شده است که مطالب کتاب و لوح فشرده همراه کتاب به نحوی تنظیم شود که علاقه‌مندان به فراگیری روش‌های تجزیه و تحلیل RNA-seq با صرف حداقل زمان، اصول این روش‌ها را دریابند و بتوانند از آن‌ها در پروژه‌های پژوهشی خود استفاده کنند.

مزیت عمده این کتاب این است که ابزارهایی که در آن معرفی می‌شوند ابزارهای استاندارد، رایگان و با متن باز هستند. استفاده از این ابزارها به پرداخت هزینه و یا عدول از اصول اخلاقی مربوط به حقوق مالکیت معنوی نیاز ندارد. چنانچه تجزیه و تحلیل با این ابزارها فراگرفته شود، استفاده از ابزارهای جدیدتری که به سرعت در حال توسعه و توزیع در جامعه علمی هستند، بسیار ساده‌تر خواهد بود و پژوهشگر همواره به جدیدترین ابزارهای تجزیه و تحلیل RNA-seq دسترسی خواهد داشت.

هرچند تلاش زیادی شده است که اشکالات نگارشی و علمی در متن وجود نداشته باشد، با این حال سپاسگزار خواهم بود چنانچه خوانندگان محترم اشکالات احتمالی و یا پیشنهادهای سازنده خود را برای ارتقای کیفی کتاب به نحو مقتضی به این جانب منعکس کنند.

علیرضا سیفی

زمستان ۱۳۹۹

www.ketab.ir