



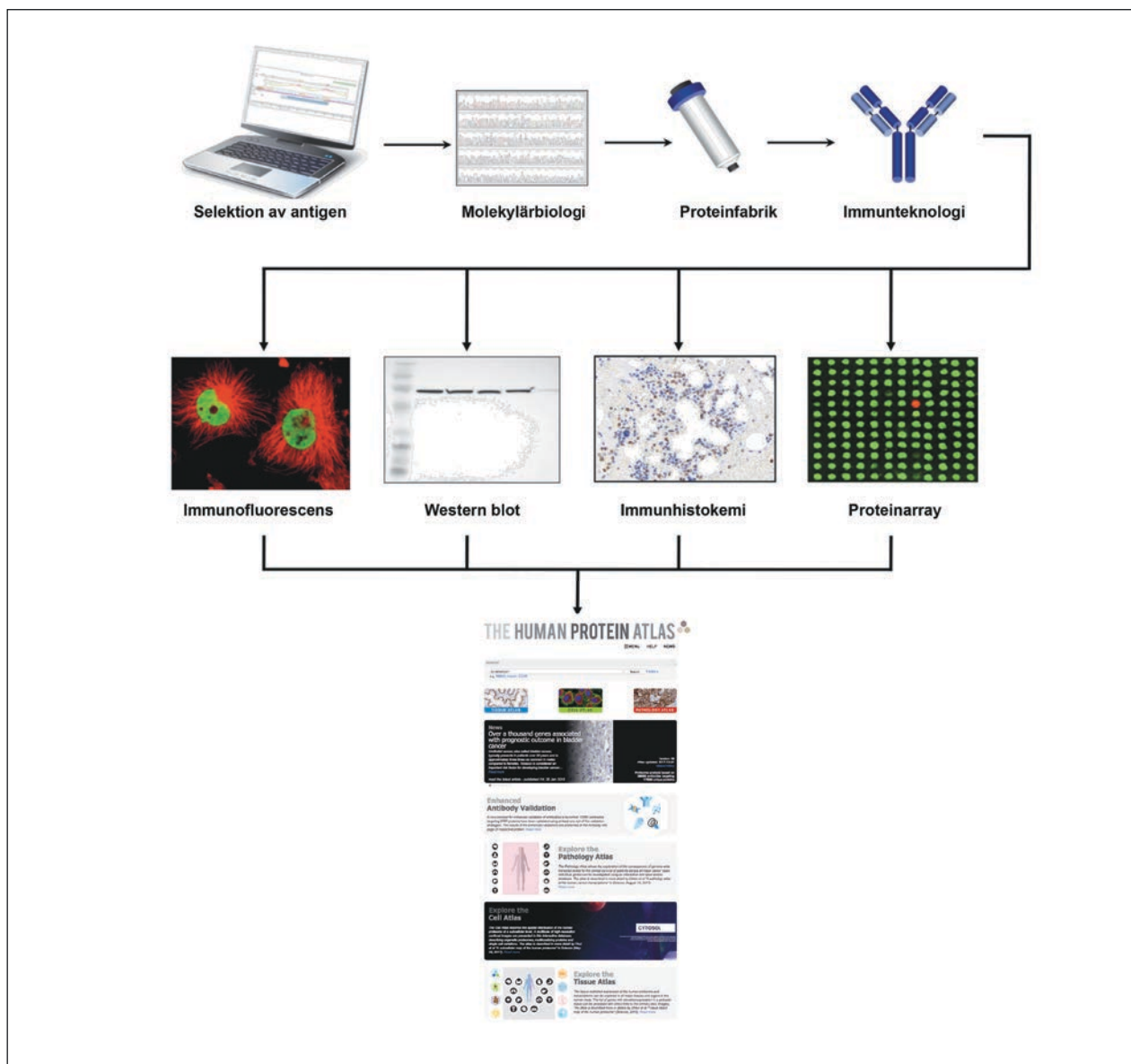
Human Protein Atlas

– kartläggningen av människans proteiner

Human Protein Atlas består av tre olika delar – en normalvävnadsatlas, en cellatlas och en patologiatlas – och innehåller data från mer än 26 000 antikroppar riktade mot 17 000 unika proteiner. När den senaste delen, patologiatlasen, presenterades hösten 2017 sågs den som en unik satsning som innebär att drömmen om individbaserad cancerbehandling tar ett stort steg framåt. Här beskriver professor **Fredrik Pontén** arbetet med kartläggningen och de framgångsfaktorer som fått det gigantiska projektet att fungera så friktionsfritt och bli en av världens mest använda biomedicinska kunskapsbaser.



Human Protein Atlas (HPA) – projektet för att kartlägga människans alla proteiner – startade 2003. Målet för HPA var att kartlägga människans alla proteiner. Bakgrunden och en förutsättning för proteinatlasprojektet var att det mänskliga genomet kartlagts bara ett par år tidigare (2001). Efter att kartan över våra gener publicerats blev det ett naturligt nästa steg att ta reda på vilka proteiner våra gener kodar för och var dessa uttrycks. Innan själva projektet startade och fick finansiering gjorde vi ett begränsat “Chromosome 21 project” och när vi såg att den tillgängliga gensekvensen gick att använda som templat för att göra rekombinanta proteiner som sedan kunde användas som antigen för framställning av specifika antikroppar, förstod vi att detta kunde skalas upp för att kartlägga alla våra proteiner. Lyckligtvis förstod även insatta personer på Knut och Alice Wallenbergs (KAW) forskningsstiftelse också detta och såg en stor potential i att Sverige kunde vara med och leda den framtida kartläggningen av människans proteiner. Vi kallade denna strategi för antikropps-baserad proteomik¹. The Human Protein Atlas-projektet möjliggjordes genom ett mycket stort anslag från KAW till professor Mathias Uhlén.



Figur 1. Flödesschema över den storskaliga produktion av antigen, antikroppar och proteinprofiler som utgör en central del av Human Protein Atlas.

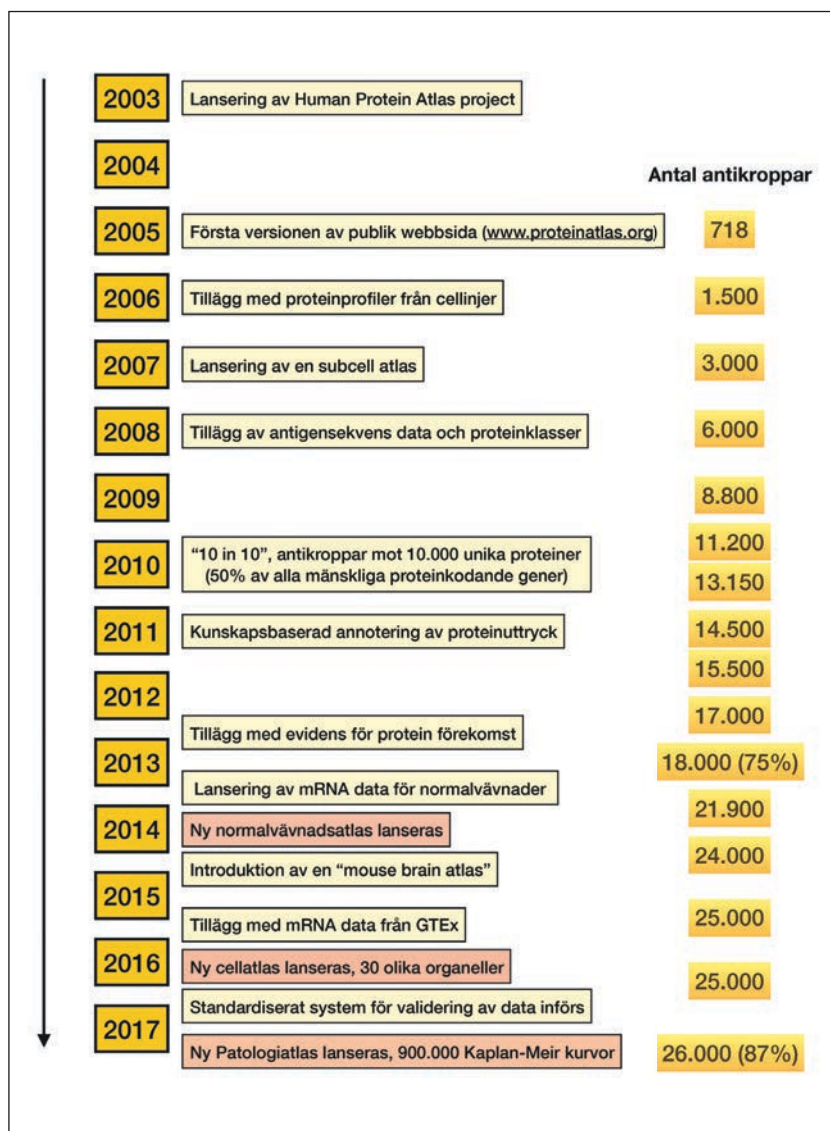
KARTA SOM VILAR PÅ TRE DELAR

Den vision som låg bakom proteinatlasprojektet var att skapa en karta över människans alla proteiner. Den metod vi utvecklade var att generera specifika antikroppar mot alla mänskliga proteiner och sedan använda dessa för att karakterisera motsvarande proteiner. Senare tillkom även målet att använda både antikroppar och de data vi fått fram genom kartläggningen till att försöka utveckla cancerbiomarkörer. När det gällde själva proteinkartan vilar den idag på tre olika delar: 1) hur uttrycks våra gener och motsvarande proteiner i normala organ och vävnader, 2) var i cellen uttrycks våra proteiner och 3) hur

uttrycks våra gener i cancer och hur relaterar detta till patientöverlevnad.

För att kunna skapa förutsättningar för storskalig tillverkning av funktionella antikroppar och för att kunna använda dessa till att immunfärga vävnader och celler, satte vi upp en "Fordfabriksliknande" modell. Vi definierade olika moduler (Figur 1) och bestämde vad varje modul månadsvis skulle leverera till nästa modul. För att hålla ihop projektets olika delar byggde vi ett gen-baserat LIMS-system där resultaten från de olika delarna kunde sammanställas och där data kunde extraheeras för publicering på nätet. Processen för varje protein initierades i en upp-

ströms bioinformatik-modul. Här selekterades unika aminosyresekvenser från varje proteinkodande gen för att kunna tillverka ett unikt rekombinant proteinfragment för att representera varje enskilt protein. Dessa proteinfragment användes sedan för immunisering och affinitetsrening av antisera så att vi i slutänden fick protein-specifika polyclonala affinitetsrenade antikroppar. Dessa antikroppar testas på proteinarrayer för specifik bindning till korrekt antigen och godkänns då för vidare testning i Western blot, immunhistokemi och immunfluorescens.



Figur 2. Tidslinje över utvecklingen av Human Protein Atlas. Hitintills har vi lanserat 18 versioner av proteinatlasen med tillkomst av data från nya antikroppar och med nya funktioner. Idag finns data för nära 90 procent av alla människans proteinkodande gener.

En viktig grundregel har alltid varit att alla de resultat som genereras inom Human Protein Atlas ska komma hela forskarsamhället till del genom att alla data ska publiceras för att bli fritt tillgängliga. För att möjliggöra detta skapade vi en webportal för proteinatlasen (www.proteinatlas.org), där vi åtminstone en gång årligen uppdaterar med både mer data från fler gener och nya funktioner. Den första versionen av proteinatlasen lanserades 2005 och innehöll proteinuttrycksdata från 718 antikroppar.

NORMALATLAS

Förutom att den humana gensekvensen fanns fritt tillgänglig hade det i slutet av 1990-talet utvecklats metoder och instrument för tillverkning av vävnadsarrayer. Att utveckla och bygga vävnadsarrayer var en förutsättning för att kunna analysera en stor mängd olika vävnader och tumörer. Vi bestämde oss tidigt för att satsa på att kartlägga hur och var alla våra proteiner uttrycks i normala organ och vävnader, då uttrycksmönstret för varje protein ger en ledtråd om funktion men också bidrar till en mer generell förståelse för vilka proteiner som är cell- eller vävnadstypspecifika och vilka proteiner som uttrycks i alla celler. Förutom att analysera normala vävnader ville vi också fokusera på cancer för att kartlägga hur genuttryck skiljer sig i cancer jämfört med normalt. Efter godkända etikansökningar fick vi tillgång till patologens arkivmaterial och använde oss av

VÄLFUNGERANDE ORGANISATION

Något som var mycket viktigt i början var den styrelse som ledde vårt arbete. Den bestod av kunniga personer från Wallenbergstiftelsen, näringslivet och akademien. Genom kvartalsvis rapportering och diskussioner kring mål och budget så fungerade detta gigantiska projekt fantastiskt väl. En nyckelperson och förutsättning för detta var förstås Mathias Uhlén som var programdirektör och operativ chef med förmåga att driva projektet framåt. Under de 15 år som projektet pågått så har vi kontinuerligt genererat data från allt fler antikroppar, men också utvecklat nya funktioner och ny kunskap (Figur 2).

"Den metod vi utvecklade var att generera specifika antikroppar mot alla mänskliga proteiner och sedan använda dessa för att karakterisera motsvarande proteiner. Senare tillkom även målet att använda både antikroppar och de data vi fått fram genom kartläggningen till att försöka utveckla cancerbiomarkörer."

överbliven vävnad från den kliniska diagnostiken och framställde paraffinklossar med vävnadsbitar från anonymiserade patienter. Utifrån dessa klossar kunde vi sedan tillverka vävnadsarrayer och för att få stor täckning av människans olika organ och vävnadstyper skapade vi vävnadsarrayer med 144 bitar normal vävnad representerande 45 olika vävnadstyper.

Snitt från dessa vävnadsarrayer används sedan för att undersöka proteinuttrycksmönstret med hjälp av immunhistokemi. Immunhistokemiska färgningar av vävnadssnitt innebär att man kan visualisera var och i vilka celler ett specifikt protein uttrycks genom att den mikroskopiska bilden finns bevarad som bakgrund för proteinuttrycksdata. Metoden är inte i sig kvantitativ men visar på relativa proteinuttrycksnivåer och ger framförallt en unik spatial information med cellulär upplösning. Efter immunhistokemiska färgningar digitaliseras resultaten genom att de immunfärgade glasen skannas för att skapa högupplösta bilder. För att sedan kunna bedöma och tolka proteinuttrycket på vävnadsnivå krävs patologisk och histologisk kompetens. För att kunna bedöma de enorma mängder immunhistokemiskt färgade vävnader som vi producerade, startade vi ett samarbete med en grupp patologer i Indien. År 2006 byggde vi upp en proteinatlas-site i Mumbai och utrustade dessa lokaler med ett högkapacitets-internet, moderna datorer och högupplösta bildskärmar. Vi byggde vidare på vårt LIMS-system och utvecklade en web-baserad modul för annotering av immunhistokemiskt färgade vävnadssnitt. Sammanlagt har närmare 12 miljoner högupplösta bilder på normalvävnader och cancer bedömts och annoterats under drygt sex års tid. Dessa annoteringsdata utgör en viktig del av proteinatlasen.

För drygt fem år sedan hade tekniker utvecklats för högkvalitativ sekvensering av transkriptom, så att man utifrån frysta vävnadsbitar också kunde kartlägga vilka mRNA-molekyler som uttrycktes i de celler som fanns i vävnaden. Då mängden av ett specifikt mRNA korrelerar till mängden av motsvarande protein, insåg vi att sekvensering av alla mRNA-molekyler i en väv-

nad skulle ge viktig och kvantitativ information beträffande genuttryck och skulle utgöra ett perfekt komplement till den kvalitativa och spatiala information som vi fick från immunhistokemi. Från Uppsala Biobank valde vi ut normalvävnader för att representera kroppens alla organ och vävnadstyper och körde sedan mRNA-sekvensering på alla dessa. År 2015 hade vi sammanställt resultaten, lanserat en alldeles ny uppdaterad version av proteinatlasen, nu med genuttrycksdata från både protein och mRNA. Allt detta gjorde att vi kunde lansera ett första utkast på en Human Protein Atlas med en karta som visade hur alla våra gener uttrycks i våra olika normala vävnader och organ. Arbetet med att integrera genuttryck på mRNA- och proteinnivå medgav ny kunskap kring hur stor del av våra gener som är vävnadsspecifika och som potentiellt kan vara mycket viktiga

både data från allt fler proteiner och nya funktioner. I cellatlasen definieras 32 olika organeller eller "compartments" där proteiner uttrycks och på liknande sätt som i normal- och canceratlasen, bedöms/annoteras här de immunfluorescens-färgade bilderna manuellt och utgör grunden för de data som publiceras i cellatlasen. Även inom arbetet med cellatlasen används data från RNA-sekvensering. Utifrån de nivåer av specifika mRNA som finns i olika cellinjer för varje gen, väljs lämpliga cellinjer ut för att undersöka var i cellen det motsvarande proteinet uttrycks. In vitro-odlade celler lämpar sig även bra för att studera när i cellcykeln vissa proteiner uttrycks och vidare kan cellinjer med fördel användas för att validera att antikroppar binder till de proteiner som de ska. År 2017 sammanfattades arbetet med att lokalisera var i cellen människans alla proteiner uttrycks och det led-

"Utöver att analysera normala vävnader ville vi också fokusera på cancer för att kartlägga hur genuttryck skiljer sig i cancer jämfört med normalt."

både som biomarkörer och som specifika targets för läkemedel. Analyserna från allt detta arbete sammanfattades och ledde till publiceringen av "Tissue based map of the human proteome", som översiktligt beskriver hur våra gener uttrycks i kroppen².

CELLATLAS

Då viktig information om ett protein kan länkas till var och hur det uttrycks bestämde vi oss tidigt för att vi skulle analysera var i cellen alla människans proteiner uttrycks. Genom att använda odlade cancercellinjer tillsammans med immunfluorescens och konfokalmikroskopi kunde vi även analysera i vilka organeller våra olika proteiner uttrycks. Den första versionen av en cellatlas lanserades redan 2007 och sedan dess har cellatlasen årligen uppdaterats med

de till publiceringen av "A subcellular map of the human proteome"³.

PATOLOGIATLAS

Utöver att analysera normala vävnader ville vi också fokusera på cancer för att kartlägga hur genuttryck skiljer sig i cancer jämfört med normalt. Efter godkända etikansökningar fick vi tillgång till patologens arkivmaterial och använde oss av överbliven vävnad från den kliniska diagnostiken och framställde paraffinklossar med vävnadsbitar från anonymiserade patienter. För att analysera proteinuttrycket i cancer gjordes motsvarande vävnadsarrayer från 216 olika tumörer representerande de 20 vanligaste formerna av mänsklig cancer och cirka 50 olika cancercellinjer. Från de allra flesta av dessa cancerformer analyseras tumörer från 12 oli-

ka patienter vilket medger en viss inblick i hur varje protein uttrycks i en viss form av cancer. Då cancerproverna var anonymiserade kunde vi däremot inte analysera om hur olika proteinuttryck eventuellt korrelerar till tumörens biologi eller övriga patientrelaterade data såsom överlevnad. Då arbetet med att analysera mRNA-uttryck i normala vävnader visat sig vara så väldigt framgångsrikt funderade vi på om man inte också för cancer kunde utnyttja de kvantitativa metoder för transkriptomanalyser som mRNA-sekvensering innebär. Projektet The Cancer Genome Atlas (TCGA) lanserades 2005 och som en del i detta gigantiska projekt utfördes mRNA-sekvensering av tumörvävnad från ett stort antal cancerpatienter. All denna data var fritt tillgänglig och nedladdningsbar, så tillsammans med ett gäng begåvade bioinformatiker beslutade vi oss för att kartlägga genuttrycket på mRNA-nivå i cancer och analysera hur olika genuttrycksnivåer korrelerade med patientöverlevnad. Detta blev ett gigantiskt bioinformatiskt projekt som år 2017 ledde fram till publiceringen av "A pathology atlas of the human cancer transcriptome"²⁴.

Till patologi Atlanten valde vi att inkludera alla de former av cancer i TCGA-databasen där det fanns mRNA-uttrycksdata från minst 100 patienter med överlevnadsdata. Det visade sig att det blev 17 olika former av cancer och data från ungefär 8 000 olika patienter. Den bioinformatiska analysen inkluderade en statistisk jämförelse mellan genuttrycksnivåer av varje gen i varje cancer. Det visade sig att uttrycksnivån för en mycket stor andel av våra gener uppvisar en signifikant korrelation med hur det går för en patient med cancer. De gener vars nivå signifikant korrelerar med patientöverlevnad benämns som prognostiska och det visar sig att det finns ungefär lika många "bra prognos-gener" som "dålig prognos-gener". En "bra prognos-gen" definieras genom att högt uttryck av denna gen är associerat med lång överlevnadstid hos patienten i motsats till en "dålig prognos-gen" där högt uttryck istället är associerat till en kortare överlevnadstid. Intressant nog finns det inga gener som är generella, så till-

vida att de signalerar bra eller dålig prognos i alla olika former av cancer, utan det är snarare så att varje cancerform har sin egen uppsättning av bra och dålig prognos-gener. Allmänt kan man dock konstatera att funktioner som relaterar till cellcyckelkontroll, proliferation och celltillväxt är överrepresenterade bland "dålig prognos-gener" och funktioner som kan kopplas

intresse för cancerforskningen och kan utgöra byggstenar i projekt som syftar till att identifiera nya kliniskt betydelsefulla cancerbiomarkörer eller till att utveckla nya specifika cancerläkemedel. Även om den nya patologi Atlanten utgör en modell för att påvisa hur olika genuttrycksnivåer kan påverka utfallet för en patient med cancer, men det är viktigt att påpeka att dessa data reflek-

"Alla dessa data som är fritt tillgängliga i patologi Atlanten är av potentiellt intresse för cancerforskningen och kan utgöra byggstenar i projekt som syftar till att identifiera nya kliniskt betydelsefulla cancerbiomarkörer eller till att utveckla nya specifika cancerläkemedel."

ihop med celldifferentiering är överrepresenterade bland "bra prognos-gener".

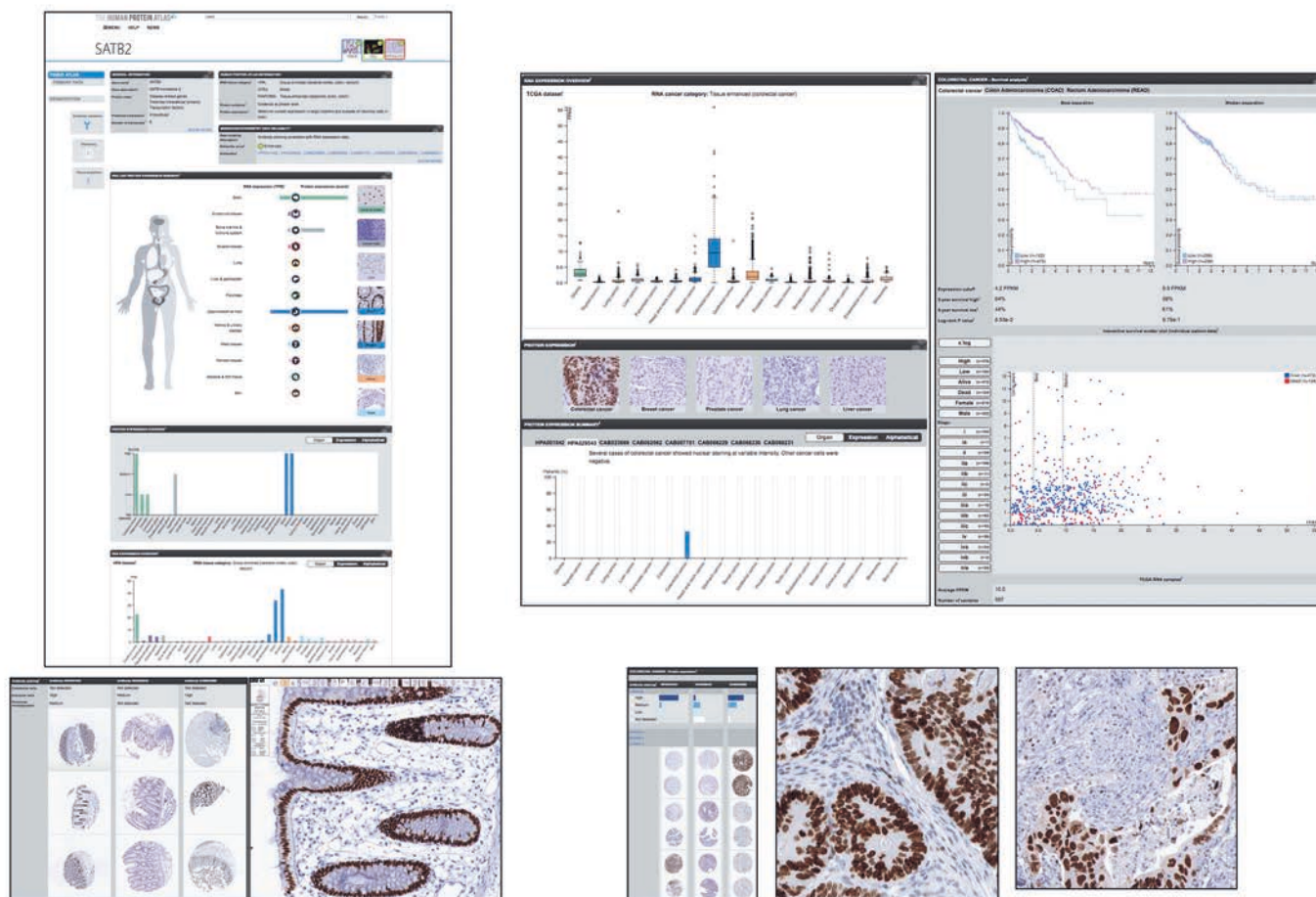
En enorm mängd (>900 000) överlevnadsanalyser, så kallade Kaplan-Meier-kurvor genererades för att kunna analysera alla gener i alla de olika inkluderade formerna av cancer. Under detta arbetet utvecklade vi ett alternativt instrument för att visualisera de data som ligger bakom en Kaplan-Meier-analys. Dessa benämns Interactive Survival Scatter plots och visar genuttrycksnivån för varje patient samtidigt som man ser hur länge varje patient överlevt sedan diagnosen ställdes. Detta ger på ett sätt en bättre översikt av alla grunddata som ligger bakom en Kaplan-Meier-kurva. I den nya patologi Atlanten visas alla dessa Interactive Survival Scatter plots för varje gen och för alla dessa 17 former av cancer. Att diagrammen benämns som "interactive" beror på att användaren själv kan filtrera vilka patienter, utifrån till exempel ett visst tumörstadium, som ska visas.

Alla dessa data som är fritt tillgängliga i patologi Atlanten är av potentiellt

terar de patientkohorter som insamlats av Cancer Genome Atlas-initiativet och på intet sätt utgör populationsbaserade kohorter eller på annat vis epidemiologiskt definierade patientstudier. För att kunna dra några verkliga slutsatser krävs att de potentiella prognosgener som visas i patologi Atlanten verifieras och valideras i oberoende väl definierade patientkohorter. Figur 3 visar resultat för SATB2-genen för att ge några exempel på en del av de gencentriska data som finns för alla proteiner.

SAMMANFATTNING

Idag består proteinatlasen (Human Protein Atlas version 18) av tre olika delar, en normalvävnadsatlas, en cellatlas och en patologi atlas och innehåller data från mer än 26 000 antikroppar riktade mot 17 000 unika proteiner. En grundsten för proteinatlasen är bilder av hög kvalitet, såväl immunhistokemiskt färgade normal- och cancervävnader, som immunfluorescensfärgade celler. Normalvävnadsatlasen beskriver genuttryck på protein- och mRNA-nivå i människokroppens alla vanliga organ



Figur 3. Exempel på data för SATB2. Uppe till vänster ses en sammanfattning av genuttryck på både mRNA- och proteinnivå i normalvävnader som visar på specifikt uttryck i kolon och rektum. Till höger visas hur SATB2 uttrycks i cancer med mRNA-data för SATB2 från 597 patienter med kolorektalcancer. Vidhängande Kaplan-Meier-kurva samt samma data visualiserad i vår introducerade interactive Survival Scatter plot. Nedtill i bilden visas exempel på SATB2 proteinuttryck i normal kolon och koloncancer med immunhistokemi.

och vävnader, och bygger på både egna data och data importerade från andra stora databaser (GTEx och FANTOM5). Cellatlasen beskriver var i cellen alla våra proteiner uttrycks och ger spatial information på subcellulär nivå. Patologi atlasen beskriver vilka gener som är prognostiska vid olika former av cancer och bygger på data från Cancer Genome Atlas-databasen, samt visar på exempel på hur alla våra proteiner uttrycks i olika former av cancer. Utöver dessa tre grundpelare finns även en stor mängd data om alla våra gener och proteiner som sammanfattas i innehållsrika kunskapsbaserade kapitel om det humana proteomet, och det finns även kapitel för utbildning och undervisning i form av en "Dictionary".

Då en stor del av proteinuttrycksdata är beroende på kvaliteten av använda

antikroppar, finns ett mycket stort fokus på antikroppsvalidering. Ambitionsnivån är mycket hög för att i slutändan kunna presentera data som bygger på validerade och verifierade metoder och reagenser. Detta är förstås av allra yttersta vikt då proteinatlasen idag utgör en grundsten för biomedicinsk forskning runt hela världen. The Human Protein Atlas (www.proteinatlas.org) har idag fler än 200 000 besökare varje månad och utgör därmed en av världens mest använda biomedicinska kunskapsbaser. Utöver fortsatt arbete med att lägga till nya gener och att höja kvaliteten på de data som presenteras för varje årlig ny version, så tillkommer även i framtiden nya kunskapsbaserade kapitel och resurser.

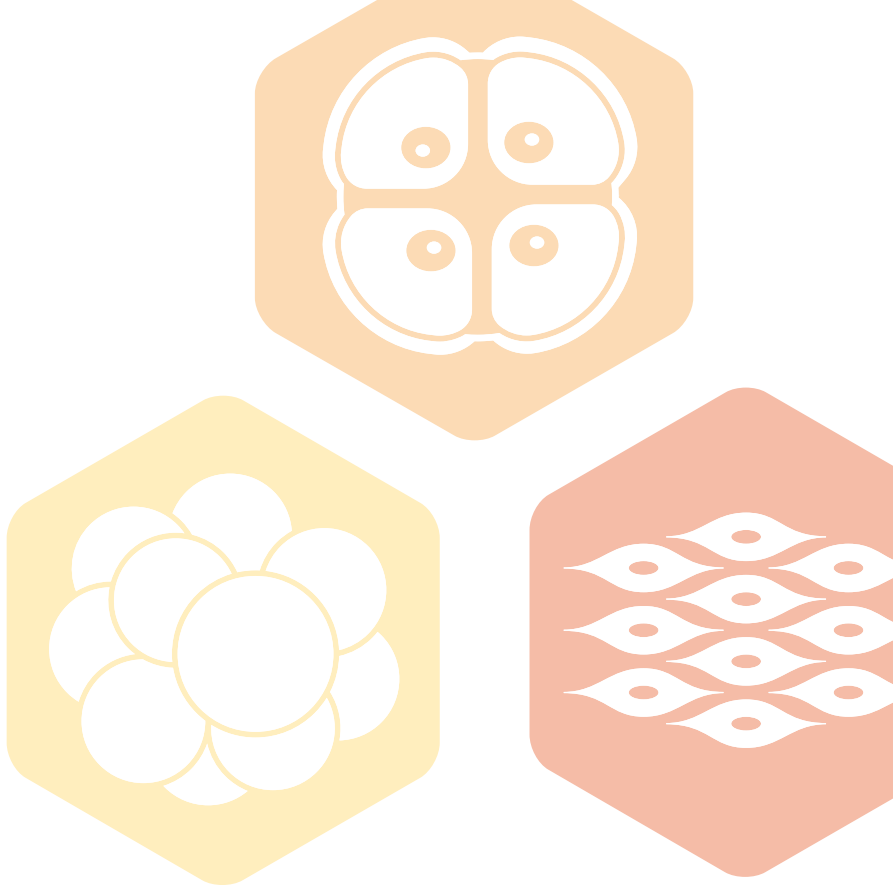
REFERENSER

1. Antibody-based proteomics for human tissue profiling. Uhlen M, Pontén F. *Mol Cell Proteomics*. 2005 Apr;4(4):384-93
2. Tissue-based map of the human proteome. Uhlen M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szegedy CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Pontén F. *Science*. 2015 Jan 23;347(6220):1260419

3. A subcellular map of the human proteome. Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, Alm T, Asplund A, Björk L, Breckels LM, Bäckström A, Danielsson F, Fagerberg L, Fall J, Gatto L, Gnann C, Hober S, Hjelmare M, Johansson F, Lee S, Lindskog C, Mulder J, Mulvey CM, Nilsson P, Oksvold P, Rockberg J, Schutten R, Schwenk JM, Siwertsson Å, Sjöstedt E, Skogs M, Stadler C, Sullivan DP, Tegel H, Winsnes C, Zhang C, Zwahlen M, Mardinoglu A, Pontén F, von Feilitzen K, Lilley KS, Uhlén M, Lundberg E. Science. 2017 May 26;356(6340)

4. A pathology atlas of the human cancer transcriptome. Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhori G, Benfeitas R, Arif M, Liu Z, Edfors F, Sanli K, von Feilitzen K, Oksvold P, Lundberg E, Hober S, Nilsson P, Mattsson J, Schwenk JM, Brunnström H, Glimelius B, Sjöblom T, Edqvist PH, Djureinovic D, Micke P, Lindskog C, Mardinoglu A, Ponten F. Science. 2017 Aug 18;357(6352).

5. A tool to facilitate clinical biomarker studies--a tissue dictionary based on the Human Protein Atlas. Kampf C, Bergman J, Oksvold P, Asplund A, Navani S, Wiking M, Lundberg E, Uhlén M, Ponten F. BMC Med. 2012 Sep 12;10:103.



FREDRIK PONTÉN, PROFESSOR, INSTITUTIONEN FÖR IMMUNOLOGI, GENETIK OCH PATOLOGI, UPPSALA UNIVERSITET, FREDRIK.PONTEN@IGP.UU.SE



- ▶ **Kostnadsfritt behandlingsstöd**
- ▶ **Enkelt** – samlingsplattform med fokus på instruktionsfilmer
- ▶ Även som **APP** – notiser för nypubliceringar

INSTRUKTIONSFILMER INOM ONKOLOGI FÖR DIG OCH DINA PATIENTER

Ett enda klick för rätt och säker läkemedelsanvändning

Beställ kostnadsfria påminnelsekort: info@medicininstruktioner.se

